

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 13, 2003, 14:45:15 ; Search time 15.7768 Seconds  
(without alignments)  
1713.885 Million cell updates/sec

Title: US-09-854-356-6

Perfect score: 5078  
Sequence: 1 MEALALCRWGLLALLPGA.....TFKGTPTAENPEYLGVDPV 919

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*

1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PCCTS.COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/Dackfilest1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4900	96.5	1255	2	US-08-625-101-2
2	4900	96.5	1255	2	US-08-336-786-2
3	4891	96.3	1255	1	US-08-467-083-68
4	4891	96.3	1255	1	US-08-414-417B-68
5	4891	96.3	1255	2	US-08-486-348A-68
6	4891	96.3	1255	2	US-08-468-545B-68
7	4891	96.3	1255	3	US-08-466-680B-68
8	4818	94.9	1255	2	US-08-484-438-8
9	3632	71.5	782	2	US-09-146-283-4
10	3632	71.5	782	3	US-08-579-823A-4
11	3632	71.5	782	4	US-09-344-195-4
12	3473	68.4	624	4	US-08-422-108-1
13	3473	68.4	624	4	US-08-422-108-1
14	1878	37.0	419	4	US-09-630-155-2
15	1645	32.4	1210	2	US-08-484-438-7
16	1645	32.4	1210	2	US-08-484-438-7
17	1533	30.2	644	1	US-08-475-035-4
18	1528.5	30.1	1308	2	US-08-336-708A-9
19	1474	29.0	580	1	US-08-414-417B-69
20	1474	29.0	580	2	US-08-486-348A-69
21	1474	29.0	580	2	US-08-486-348A-69
22	1474	29.0	580	3	US-08-486-348A-69
23	1466.5	28.9	911	2	US-08-484-438-10
24	1451	28.3	1058	2	US-08-484-438-4
25	1435	28.3	1342	1	US-07-978-895-4
26	1435	28.3	1342	1	US-08-484-438-9
27	1435	28.3	1342	2	US-08-473-119-4

28	1435	28.3	1342	2	US-08-475-352-4	Sequence 4, Appl1
29	1427.5	28.1	1343	6	5183884-4	Patent No. 5183884
30	1028.5	20.3	478	4	US-09-570-454-2	Sequence 2, Appl1
31	493	9.7	97	1	US-08-421-356-3	Sequence 3, Appl1
32	493	9.7	97	4	US-09-046-783-3	Sequence 3, Appl1
33	368	7.2	1382	2	US-08-737-715-2	Sequence 2, Appl1
34	368	7.2	1382	4	US-09-457-040B-7	Sequence 2, Appl1
35	313.5	6.2	1367	2	US-08-625-819-2	Sequence 2, Appl1
36	311.5	6.1	516	3	US-08-746-559A-4	Sequence 2, Appl1
37	311.5	6.1	1367	2	US-08-249-687C-2	Sequence 2, Appl1
38	311.5	6.1	1367	3	US-08-746-559A-2	Sequence 2, Appl1
39	305.5	6.0	705	2	US-08-456-647B-4	Sequence 2, Appl1
40	305.5	6.0	705	2	US-08-237-401A-4	Sequence 2, Appl1
41	303.5	6.0	1367	4	US-08-864-641B-18	Sequence 4, Appl1
42	295.5	5.8	486	3	US-08-746-559A-5	Sequence 18, Appl1
43	291	5.7	370	4	US-08-857-076-104	Sequence 5, Appl1
44	269.5	5.3	541	2	US-08-484-438-6	Sequence 104, App
45	267.5	5.3	1724	4	US-08-857-076-12	Sequence 6, Appl1

## ALIGNMENTS

```
RESULT 1
US-08-625-101-2
; Sequence 2, Application US/08625101
; Patent No. 5869445
; GENERAL INFORMATION:
; APPLICANT: Cheever, Martin A.
; TITLE OF INVENTION: COMPOUNDS FOR ELICITING OR ENHANCING IMMUNE
; TITLE OF INVENTION: REACTIVITY TO HER-2/neu PROTEIN FOR PREVENTION
; TITLE OF INVENTION: OR TREATMENT OF MALIGNANCIES IN WHICH THE HER-2/neu
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/625,101
; FILING DATE: 01-APR-1996
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Sharkkey, Richard G.
; REGISTRATION NUMBER: 32,629
; REFERENCE/DOCKET NUMBER: 920010.448C7
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1255 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-625-101-2

Query Match 96.5%; Score 4900; DB 2; Length 1255;
Best Local Similarity 73.2%; Pred. No. 0;
Matches 919; Conservative 0; Mismatches 0; Indels 336; Gaps 1;

CY 1 MEALALCRWGLLALLPGAASFOVCTGDMKRLRASPETHLDMRLHYGCGQVVGNTL 60
DB 1 MEALALCRWGLLALLPGAASFOVCTGDMKRLRASPETHLDMRLHYGCGQVVGNTL 60
```

```

QY 61 ELTYLPTNASLSFLQDIOEVQGVYLIAHNOVRYPLQRLRIYRGTOLEFEDNVALAVDNG 120
DB 61 ELTYLPTNASLSFLQDIOEVQGVYLIAHNOVRYPLQRLRIYRGTOLEFEDNVALAVDNG 120
QY 121 DPLNNTPTVPGASPGGLRELOLRSLTEILKGVLIQRPOLCYODPTILMKDIFKKNOLA 180
DB 121 DPLNNTPTVPGASPGGLRELOLRSLTEILKGVLIQRPOLCYODPTILMKDIFKKNOLA 180
QY 181 LTLIDNRSRACHPCSPMCKSGRCWGESSEDCOSLTRVYVAGGACAKGKPLPTDCCHEQC 240
DB 181 LTLIDNRSRACHPCSPMCKSGRCWGESSEDCOSLTRVYVAGGACAKGKPLPTDCCHEQC 240
QY 241 AACCTGPKHSDCLACILHFNHSGICEILHCPALVYNTDTEFSMNPBGRYTFGASCTYACP 300
DB 241 AACCTGPKHSDCLACILHFNHSGICEILHCPALVYNTDTEFSMNPBGRYTFGASCTYACP 300
QY 301 YNLTSDVGSCTLVCPPLHNOEVAEDGTQRCCKSKPCARVICYGLMEHLREYRAVTSAN 360
DB 301 YNLTSDVGSCTLVCPPLHNOEVAEDGTQRCCKSKPCARVICYGLMEHLREYRAVTSAN 360
QY 361 IOEFAGCKITFGSLATLPESFDGDPASNTAPLOPELOVETLEETGYLYISAMPDLP 420
DB 361 IOEFAGCKITFGSLATLPESFDGDPASNTAPLOPELOVETLEETGYLYISAMPDLP 420
QY 421 DLSVEONLQVIRGRILHNGAVSLTLOGLSIMLGRSLRELSGSLALIHNTHLCEVHTV 480
DB 421 DLSVEONLQVIRGRILHNGAVSLTLOGLSIMLGRSLRELSGSLALIHNTHLCEVHTV 480
QY 481 PMDOLFNNPQALHTANRPEDCEVGEGLAC HOLCARGHCGWPGPTOCVNC SOFLRGQEC 540
DB 481 PMDOLFNNPQALHTANRPEDCEVGEGLAC HOLCARGHCGWPGPTOCVNC SOFLRGQEC 540
QY 541 VEECRVIOGLPREVYNAHCLPCHECOQONGSTTCGPRADQVCAAHKDPPEFCVARC 600
DB 541 VEECRVIOGLPREVYNAHCLPCHECOQONGSTTCGPRADQVCAAHKDPPEFCVARC 600
QY 601 PSQVPRDLSYMPIMKFPDEBACOPCPLNCTHSCVDLDDKCPAEORASPLTS----- 653
DB 601 PSQVPRDLSYMPIMKFPDEBACOPCPLNCTHSCVDLDDKCPAEORASPLTS----- 653
QY 654 ----- 653
DB 654 ----- 653
QY 661 ILLVVVLGVFGLIKRRQOKIRKTYMRLLQETELVEPLTPSGAMPNOAMRIKETEL 720
DB 661 ILLVVVLGVFGLIKRRQOKIRKTYMRLLQETELVEPLTPSGAMPNOAMRIKETEL 720
QY 654 ----- 653
DB 654 ----- 653
QY 721 RKVYVLSGAGATYKGIWIPDGENVKIPVAKVIRENTSPKANEILDEAYVNAVGGSP 780
DB 721 RKVYVLSGAGATYKGIWIPDGENVKIPVAKVIRENTSPKANEILDEAYVNAVGGSP 780
QY 654 ----- 653
DB 654 ----- 653
QY 781 YVSRLLGICLSTVQVLTQMLPYGCLLDHVENRGRGLSGQDLNKMOMIAKMSLYLEDVR 840
DB 781 YVSRLLGICLSTVQVLTQMLPYGCLLDHVENRGRGLSGQDLNKMOMIAKMSLYLEDVR 840
QY 654 ----- 653
DB 654 ----- 653
QY 841 LVHRDLAARNVLYKSPNHVKITDFGLARLIDIDEYHADGKVP IKMMALESILRRFT 900
DB 841 LVHRDLAARNVLYKSPNHVKITDFGLARLIDIDEYHADGKVP IKMMALESILRRFT 900
QY 654 ----- 653
DB 654 ----- 653
QY 901 HQSDVWSYGVTVWELMTFGANPYDGIIPAREIPDLLEKGERLPORPCTIIVYMIWKCWM 960
DB 901 HQSDVWSYGVTVWELMTFGANPYDGIIPAREIPDLLEKGERLPORPCTIIVYMIWKCWM 960
QY 654 ----- 684
DB 654 ----- 684
QY 961 IDSECRPFRELFVSFMSMARDPQFVYVYQNEEDLGPASPLDSTYFRSLLEDDMDLDA 1020
DB 961 IDSECRPFRELFVSFMSMARDPQFVYVYQNEEDLGPASPLDSTYFRSLLEDDMDLDA 1020
QY 685 BEYLVPQGFPCPDAPGAGGVVHHRHSSSTRSGGDLTLGLEPSEEDAPRSLAPSEG 744
DB 685 BEYLVPQGFPCPDAPGAGGVVHHRHSSSTRSGGDLTLGLEPSEEDAPRSLAPSEG 744
QY 1021 EBYLVPOGFCPCPDAPGAGGVVHHRHSSSTRSGGDLTLGLEPSEEDAPRSLAPSEG 1080
DB 1021 EBYLVPOGFCPCPDAPGAGGVVHHRHSSSTRSGGDLTLGLEPSEEDAPRSLAPSEG 1080
QY 745 AGSDVFDGDLGMAKGIQSLPTHPSPLQRSSEPTVPLBSEMGYVAAPLTCSPQPEYV 804
DB 745 AGSDVFDGDLGMAKGIQSLPTHPSPLQRSSEPTVPLBSEMGYVAAPLTCSPQPEYV 804
QY 1081 AGSDVFDGDLGMAKGIQSLPTHPSPLQRSSEPTVPLBSEMGYVAAPLTCSPQPEYV 1140
DB 1081 AGSDVFDGDLGMAKGIQSLPTHPSPLQRSSEPTVPLBSEMGYVAAPLTCSPQPEYV 1140

```

```

QY 805 NOPDYRPPSPREGPLPAPRAPAGATLERKTLSPGKNGVKDVAFGAVENPEYLPQ 864
DB 1141 NOPDYRPPSPREGPLPAPRAPAGATLERKTLSPGKNGVKDVAFGAVENPEYLPQ 1200
QY 865 GGAAPQPPPPAFSPAFDNLTYWDDPPRRGAPPPSTFKGTPTFAENPEYLGIDVPY 919
DB 1201 GGAAPQPPPPAFSPAFDNLTYWDDPPRRGAPPPSTFKGTPTFAENPEYLGIDVPY 1255

RESULT 2
US-08-356-786-2
; Sequence 2, Application US/08356786
; Patent No. 5877305
; GENERAL INFORMATION:
; APPLICANT: Huston, James S.
; APPLICANT: Oppertmann, Hermann
; APPLICANT: Houston, L. L.
; APPLICANT: Rung, David B.
; TITLE OF INVENTION: Biosynthetic Binding Protein for Cancer
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Edmund R. Pitcher, Testa, Hurwitz, & Thibault
; STREET: Exchange Place, 53 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/356,786
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/831,967
; FILING DATE: 06-FEB-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Pitcher, Edmund R.
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: CRP-053
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ. ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1255 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-356-786-2

Query Match 96.5%; Score 4900; DB 2; Length 1255;
Best Local Similarity 73.2%; Pred. No. 0;
Matches 919; Conservative 0; Mismatches 0; Indels 336; Gaps 1;

```

Db 181 LTLIDTNRSRACHPCSPMKGSKRCWGESSESDCSLRTTVCAGGACRCKPLPTDCHEQC 240  
QY 241 AAGCTGPKHSDCIACLFHNSGICELHCPALVYNTDTFESMPNBEGRYFGASCVTACP 300  
Db 241 AAGCTGPKHSDCIACLFHNSGICELHCPALVYNTDTFESMPNBEGRYFGASCVTACP 300  
QY 301 YNTLSTDVSGCTLVCPHLNHOEYTAEDGTORCEKSKPCARVCYGLGMEHLREVRATYSAN 360  
Db 301 YNTLSTDVSGCTLVCPHLNHOEYTAEDGTORCEKSKPCARVCYGLGMEHLREVRATYSAN 360  
QY 361 IOEFACCKRIFGSLAFPESEFDGDPASNTAPLOPEOLQVETLEITGYLYISAMPDLSL 420  
Db 361 IOEFACCKRIFGSLAFPESEFDGDPASNTAPLOPEOLQVETLEITGYLYISAMPDLSL 420  
QY 421 DLSVFNQLOVYIRGRILHNGAYSLTLOGLSISMLGSLRLSGSLALIHNTHLCEVHTV 480  
Db 421 DLSVFNQLOVYIRGRILHNGAYSLTLOGLSISMLGSLRLSGSLALIHNTHLCEVHTV 480  
QY 481 PMDQFRNPQHALLHTANRPEDCEVGEGLACHOLCARGHGMGPGPQVCVNCQSLRGQEC 540  
Db 481 PMDQFRNPQHALLHTANRPEDCEVGEGLACHOLCARGHGMGPGPQVCVNCQSLRGQEC 540  
QY 541 VEECRVLOGLPREYVNAHRLCPHPCQOPONGSVTCFGEPAQCVACAHYKDPFCVARC 600  
Db 541 VEECRVLOGLPREYVNAHRLCPHPCQOPONGSVTCFGEPAQCVACAHYKDPFCVARC 600  
QY 601 PSQVNDLSMPYKRPDEGACQPCPINCTHSCVDLDKGCACABORASPLTS----- 653  
Db 601 PSQVNDLSMPYKRPDEGACQPCPINCTHSCVDLDKGCACABORASPLTS----- 653  
QY 654 ----- 653  
Db 654 ----- 653  
QY 721 RKVKVLSGAFGTGVKGIWIPDGENVKIPAIKVLRENTSPKANKELDEAYVMAGVSP 780  
Db 721 RKVKVLSGAFGTGVKGIWIPDGENVKIPAIKVLRENTSPKANKELDEAYVMAGVSP 780  
QY 781 YVSRLIGLSTVOLVTLMPYCLLDHYRENRGLSGODLLMCMQIAKMSYLEDYR 840  
Db 781 YVSRLIGLSTVOLVTLMPYCLLDHYRENRGLSGODLLMCMQIAKMSYLEDYR 840  
QY 841 LVHRLDAANVLKSPNHVKITDFGLARLLIDETRYHADGCKVPLKMALESILRRPT 900  
Db 841 LVHRLDAANVLKSPNHVKITDFGLARLLIDETRYHADGCKVPLKMALESILRRPT 900  
QY 901 HOSDVMYGVTWELMTFGAKPYDIPAREIPDLLEKGERLPQPICTIDVYIMVCKWM 960  
Db 901 HOSDVMYGVTWELMTFGAKPYDIPAREIPDLLEKGERLPQPICTIDVYIMVCKWM 960  
QY 961 IDSECRPRRELVSERMAARDQRFVYIIONEDLGASPLDSTFYSLLEDDMGDLVDA 1020  
Db 961 IDSECRPRRELVSERMAARDQRFVYIIONEDLGASPLDSTFYSLLEDDMGDLVDA 1020  
QY 1021 EETLVPOQGFPCDPAPAGAGWVHHRRSSSTRSGGDLTLGLEPSEEAERPLAPSEK 744  
Db 1021 EETLVPOQGFPCDPAPAGAGWVHHRRSSSTRSGGDLTLGLEPSEEAERPLAPSEK 744  
QY 1081 AGSDVFDGDLGMAKGLDLSLPHDPSPLQRYSEDPVPLPSETDGVYAFVFGAVENPEVLTQ 1140  
Db 1081 AGSDVFDGDLGMAKGLDLSLPHDPSPLQRYSEDPVPLPSETDGVYAFVFGAVENPEVLTQ 1140  
QY 1141 NODVAPRQSPREGLPARPAGATLERPKTISPKNGVYKDVFAFGAVENPEVLTQ 1200  
Db 1141 NODVAPRQSPREGLPARPAGATLERPKTISPKNGVYKDVFAFGAVENPEVLTQ 1200  
QY 1201 GGAAPRPHPPAFSAFNDLYYWDODPRERGAAPTSTFKGPTAENPEYGLDVPV 919  
Db 1201 GGAAPRPHPPAFSAFNDLYYWDODPRERGAAPTSTFKGPTAENPEYGLDVPV 919

RESULT 3  
US-08-467-083-68

Sequence 68, Application US/08467083  
Patent No. 5726023  
GENERAL INFORMATION:  
APPLICANT: Cheever, Martin A.  
APPLICANT: Disis, Mary L.  
TITLE OF INVENTION: IMMUNE REACTIVITY TO HER-2/NEU PROTEIN  
TITLE OF INVENTION: FOR DIAGNOSIS AND TREATMENT OF MALIGNANCIES IN WHICH THE  
NUMBER OF SEQUENCES: 68  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Seed and Berry  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: US  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/467,083  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/414,417  
FILING DATE: 06-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Sharkey, Richard G.  
REGISTRATION NUMBER: 32,629  
REFERENCE/DOCKET NUMBER: 920010.448C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
TELEX: 3723836 SEDANDBERY  
INFORMATION FOR SEQ. ID NO: 68:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1255 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-467-083-68  
Query Match 96.3%; Score 4891; DB 1; Length 1255;  
Best Local Similarity 73.1%; Pred. No. 0;  
Matches 918; Conservative 0; Mismatches 1; Indels 336; Gaps 1;  
QY 1 METALCRNGILLALALPPGAASOVCTGDMKRLPASPETHIDMLRHLXYOGQOVQGNL 60  
Db 1 METALCRNGILLALALPPGAASOVCTGDMKRLPASPETHIDMLRHLXYOGQOVQGNL 60  
QY 61 ELTYLPTNLSLFLQDIQEVQGYVLIANNOVROPVLRRLRYRGTOLEFEDNALAVLDNG 120  
Db 61 ELTYLPTNLSLFLQDIQEVQGYVLIANNOVROPVLRRLRYRGTOLEFEDNALAVLDNG 120  
QY 121 DPLNNTPTVGTASPGGLRELQRLSLREITLKGVLTIORNPOLCYODITLKKDIFHKNNOLA 180  
Db 121 DPLNNTPTVGTASPGGLRELQRLSLREITLKGVLTIORNPOLCYODITLKKDIFHKNNOLA 180  
QY 181 LTLIDTNRSRACHPCSPMKGSKRCWGESSESDCSLRTTVCAGGACRCKPLPTDCHEQC 240  
Db 181 LTLIDTNRSRACHPCSPMKGSKRCWGESSESDCSLRTTVCAGGACRCKPLPTDCHEQC 240  
QY 241 AAGCTGPKHSDCIACLFHNSGICELHCPALVYNTDTFESMPNBEGRYFGASCVTACP 300  
Db 241 AAGCTGPKHSDCIACLFHNSGICELHCPALVYNTDTFESMPNBEGRYFGASCVTACP 300  
QY 301 YNTLSTDVSGCTLVCPHLNHOEYTAEDGTORCEKSKPCARVCYGLGMEHLREVRATYSAN 360  
Db 301 YNTLSTDVSGCTLVCPHLNHOEYTAEDGTORCEKSKPCARVCYGLGMEHLREVRATYSAN 360  
QY 361 IOEFACCKRIFGSLAFPESEFDGDPASNTAPLOPEOLQVETLEITGYLYISAMPDLSL 420  
Db 361 IOEFACCKRIFGSLAFPESEFDGDPASNTAPLOPEOLQVETLEITGYLYISAMPDLSL 420

Db	361	IQPAGCKKIFGSLAFLESPFODGPNASNPAPLOPQLOQYFETLEBITGYIXISAMPDLP	420
QY	421	DLSTFQWLOYTRGRILHNGAYSLITLOGLOISWLGRLSELGSGALIHNTHLCEVHTY	480
Db	421	DLSTFQWLOYTRGRILHNGAYSLITLOGLOISWLGRLSELGSGALIHNTHLCEVHTY	480
QY	481	PMDDLFENPHQALLHTANPREDECEVGEGLACHOLCARGCMCPGTQVCNCSQFLRGEC	540
Db	481	PMDDLFENPHQALLHTANPREDECEVGEGLACHOLCARGCMCPGTQVCNCSQFLRGEC	540
QY	541	VEBCRVLOGLPREYVNAHRHCLPCHPECPQONGSVTCFGEADQCYACAHYKDPPECVARC	600
Db	541	VEBCRVLOGLPREYVNAHRHCLPCHPECPQONGSVTCFGEADQCYACAHYKDPPECVARC	600
QY	601	PSGKPLSTMPILMKPFDEEGACOPCINCTHSCVDLDDKGAPORASPLTS	653
Db	601	PSGKPLSTMPILMKPFDEEGACOPCINCTHSCVDLDDKGAPORASPLTS	653
QY	654	-----	653
Db	661	ILLVVLGVVFGILIKRROCKIRKTYMRLLJETELVEPLTPSGAMPQAOMRIKTEL	720
QY	654	-----	653
Db	721	RKVKVLGSGAFGYKGIWIPDENVKIPYAIKVLRENTSPKANKELIDEAYVNAVGVSP	780
QY	654	-----	653
Db	781	YVSRLLGICLSTIYQVLTQOLMPRGCLLDHVRKRGRLSGQDLLNMCQIANGMSTLEBYR	840
QY	654	-----	653
Db	841	LVRHDLAARVNLVKSPMNHVITDFGLARLLDIDETEYHADGKVPILKMWALESLRRFT	900
QY	654	-----	653
Db	901	HQSDVMSYGVTVWELMTFGAKPYDGIIPANEIPDLLEKGERLPOPICTIDYIMYMKWM	960
QY	654	-----QNEDLGASPLDSTFYRSLLEDDMDGLDVA	684
Db	961	IDSECRPRFRELVEFSRMARDPQRFVVIQNEDLGASPLDSTFYRSLLEDDMDGLDVA	1020
QY	685	EEYLVPOQGFPCDPAPAGAGMWHHRHSSSTRSGGDLTCLLEPSEBEAPSPPLABSEG	744
Db	1021	EEYLVPOQGFPCDPAPAGAGMWHHRHSSSTRSGGDLTCLLEPSEBEAPSPPLABSEG	1080
QY	745	AGSDVFGDGLGMAKLOSLPTHDSPIORXSEDPVLPSETDGYAPALPLCSQPPEYV	804
Db	1081	AGSDVFGDGLGMAKLOSLPTHDSPIORXSEDPVLPSETDGYAPALPLCSQPPEYV	1140
QY	805	NQDPVRQOPSPREGRPLPAARPAAGATLERKTLSPGKNGVAVDVEAFGAVENPEYLPQ	864
Db	1141	NQDPVRQOPSPREGRPLPAARPAAGATLERKTLSPGKNGVAVDVEAFGAVENPEYLPQ	1200
QY	865	GGAPQOHPPRAPSAPFDNLXYWDDOPRERGAPSTFTFKTPTAENPEYIGLDVY	919
Db	1201	GGAPQOHPPRAPSAPFDNLXYWDDOPRERGAPSTFTFKTPTAENPEYIGLDVY	1255

RESULT 4  
 US-08-414-417B-68  
 : Sequence 68 Application US/08414417B  
 : Patent No. 5801005  
 : GENERAL INFORMATION:  
 : APPLICANT: Cheever, Martin A.  
 : APPLICANT: Disis, Mary L.  
 : TITLE OF INVENTION: IMMUNE REACTIVITY TO HER-2/neu PROTEIN  
 : TITLE OF INVENTION: FOR DIAGNOSIS AND TREATMENT OF MALIGNANCIES IN WHICH THE  
 : TITLE OF INVENTION: HER-2/neu ONCOGENE IS ASSOCIATED  
 : NUMBER OF SEQUENCES: 69  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: Seed and Berry LLP  
 : STREET: 6300 Columbia Center, 701 Fifth Avenue

Query Match	Best Local Similarity	96.3%;	Score 4891;	DB 1;	Length 1255;	
Matches	918;	Conservative	0;	Mismatches	1;	Indels 336; Gaps 1;
QY 1	MELAAICRWGLLALLP	PGAASTGYCTGDMKRLRPAS	ETHLDMRLHYGCGYVQGNL	60		
DB 1	MELAAICRWGLLALLP	PGAASTGYCTGDMKRLRPAS	ETHLDMRLHYGCGYVQGNL	60		
QY 61	ELTYPTNASISFODIO	DEGVGYVLIANOVQVPLQRLRYRG	QGLFEDNVALVLNDG	120		
DB 61	ELTYPTNASISFODIO	DEGVGYVLIANOVQVPLQRLRYRG	QGLFEDNVALVLNDG	120		
QY 121	DLPLNTTGVTSAGPGL	RELOLRSLTEILKGGVLIQORNPOLCY	ODPTLMKIDIFHRNNOLA	180		
DB 121	DLPLNTTGVTSAGPGL	RELOLRSLTEILKGGVLIQORNPOLCY	ODPTLMKIDIFHRNNOLA	180		
QY 181	LTLIDTNSRACHPCSP	CMKSGRCSGEESSEDCOSLRTFYC	AGGACRCAGRLPTDCHEQC	240		
DB 181	LTLIDTNSRACHPCSP	CMKSGRCSGEESSEDCOSLRTFYC	AGGACRCAGRLPTDCHEQC	240		
QY 241	AAGCGPRHSDCLAH	FNHSGICELHCPALVTYNTDFFES	MPNDEGRYTFGASCVTACP	300		
DB 241	AAGCGPRHSDCLAH	FNHSGICELHCPALVTYNTDFFES	MPNDEGRYTFGASCVTACP	300		
QY 301	YNYLTDVSGCTLYCP	PLHNOEYTAEDGTORCEKSKPCAR	VCYGLGMEHLREYRAVTSAN	360		
DB 301	YNYLTDVSGCTLYCP	PLHNOEYTAEDGTORCEKSKPCAR	VCYGLGMEHLREYRAVTSAN	360		
QY 361	IQEFAGCKKIFGSLA	FLDESFDGDPASNTAPLQPOLO	YFETLLEBITGYLISAMPDILP	420		
DB 361	IQEFAGCKKIFGSLA	FLDESFDGDPASNTAPLQPOLO	YFETLLEBITGYLISAMPDILP	420		
QY 421	DLISVQNTQVIRGR	LHNGAYSILTQIGLISWLGRLS	RELSSGALLHNNTHLCFVRY	480		
DB 421	DLISVQNTQVIRGR	LHNGAYSILTQIGLISWLGRLS	RELSSGALLHNNTHLCFVRY	480		
QY 481	PMDDQLFRPHQALL	TANPPEDECVGEGIGLACHQ	LCLARGCMGPGPTQCNCSQPIRGQEC	540		
DB 481	PMDDQLFRPHQALL	TANPPEDECVGEGIGLACHQ	LCLARGCMGPGPTQCNCSQPIRGQEC	540		
QY 541	VEECVTLQGLPREYV	ANARHCLPCHECO	PQNSVYTCFGEDAQCVA	CAHAYKDP	PPCVARC	600
DB 541	VEECVTLQGLPREYV	ANARHCLPCHECO	PQNSVYTCFGEDAQCVA	CAHAYKDP	PPCVARC	600
QY 601	PSGVKPDLSYMDIMK	FPDEEGACQPCPINC	HSYVDLDDKCGCPADQ	RASPLTS-----		653



```

Db 601 PSGVKPDLSTYMPIMKFPDEEGACQPCPINCTHSCVDLDDKGCAPDQASPLTSTISAVVG 660
Qy 654 ----- 653
Db 661 ILLVVLGVVFGILLIRROOKIRKRYTMRLLOETELVEPLTPSGAMPNOQMRILKETEL 720
Qy 654 ----- 653
Db 721 RKVKVLGSGAFGTYYKGIWIPDGENYKIPYAIKVLRENTSPKANKELIDEAYMAGVGP 780
Qy 654 ----- 653
Db 781 YVSRLLGICLTSTVOLVTOLMPYGCLLDHVRENRGLSGODLLNMCQIAKMSYLEDR 840
Qy 654 ----- 653
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLIDETEHADGKVPITKMALESILRRFT 900
Qy 654 ----- 653
Db 901 HOSDWSYGVYTWELMTFGAKPRYDGLIPAREIPDLLEKGERLPOPICTIDYIMYKCM 960
Qy 654 ----- 653
Db 961 IDSECARPRELSEFSRMAQDQRFVYIQNEDLGASPLDSTFYRSLLDDMDGLVDA 1020
Qy 685 EEVLVQOQFPCPDPAFGAGVNHRRSSSTRSGGDLTLGLEPSEEAAPRSLAPSSG 744
Db 1021 EEVLVQOQFPCPDPAFGAGVNHRRSSSTRSGGDLTLGLEPSEEAAPRSLAPSSG 1080
Qy 745 AGSDVFDGLGMAKAGLSLPTHDSPLORYSEDPVPLPSETDGVVPLTCSPOPEY 804
Db 1081 AGSDVFDGLGMAKAGLSLPTHDSPLORYSEDPVPLPSETDGVVPLTCSPOPEY 1140
Qy 805 NOPDVRPQPPSPREGPLPAARPAATLERPKTLSPGKNGVYKDVFAFGAVENPEYLTPQ 864
Db 1141 NOPDVRPQPPSPREGPLPAARPAATLERPKTLSPGKNGVYKDVFAFGAVENPEYLTPQ 1200
Qy 865 GGAAPQHPHPAPSPAFDNLVYWDOPPERGAPSPSTFKTPTAENEYGLDVPV 919
Db 1201 GGAAPQHPHPAPSPAFDNLVYWDOPPERGAPSPSTFKTPTAENEYGLDVPV 1255

REMARKS
US-08-486-348A-68
Sequence 68, Application US/08486348A
Patent No. 3846358
GENERAL INFORMATION:
APPLICANT: Cheever, Martin A.
APPLICANT: Disis, Mary L.
TITLE OF INVENTION: IMMUNE REACTIVITY TO HER-2/neu PROTEIN
TITLE OF INVENTION: FOR DIAGNOSIS AND TREATMENT OF MALIGNANCIES IN WHICH THE
NUMBER OF SEQUENCES: 69
NUMBER OF SEQUENCES: 69
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed and Berry LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: US
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/486,348A
FILING DATE: 07-JUN-1995
CLASSIFICATION: A24
ATTORNEY/AGENT INFORMATION:
NAME: Sharkey, Richard G.
REGISTRATION NUMBER: 32,629

```

```

REFERENCE/DOCKET NUMBER: 920010.448C6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 68:
SEQUENCE CHARACTERISTICS:
LENGTH: 1255 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-486-348A-68

Query Match          96.3% Score 4891; DB 2; Length 1255;
Best Local Similarity 73.1% Pred No. 0;
Matches 918; Conservative 0; Mismatches 1; Indels 336; Gaps 1;

Qy 1 MELAALCRWGLLALLPAPGAATQVCTGDMKRLRPAETHLMDLRHLYOGQVVOGNTL 60
Db 1 MELAALCRWGLLALLPAPGAATQVCTGDMKRLRPAETHLMDLRHLYOGQVVOGNTL 60
Qy 61 ELTYLPTNASTSLFDIOIEYQGVVLAHNOVQVPLQRLRYNGTOLFEDNVALAVLDNG 120
Db 61 ELTYLPTNASTSLFDIOIEYQGVVLAHNOVQVPLQRLRYNGTOLFEDNVALAVLDNG 120
Qy 121 DPLNNTPTVGASPGGLRELOLSLFEILKGGVLIQRNPOLCYODTILMKDIFHKNNOLA 180
Db 121 DPLNNTPTVGASPGGLRELOLSLFEILKGGVLIQRNPOLCYODTILMKDIFHKNNOLA 180
Qy 181 LTLIDNRSRACHPCSPMKGSRGWSESDCOSLRTVCAGGACARCKGPLPTDCHEQC 240
Db 181 LTLIDNRSRACHPCSPMKGSRGWSESDCOSLRTVCAGGACARCKGPLPTDCHEQC 240
Qy 241 AAGCTPKHSDDLACHFNHSGICELHCPALTYNTDTFESMPNPGRTTFGASCYTACP 300
Db 241 AAGCTPKHSDDLACHFNHSGICELHCPALTYNTDTFESMPNPGRTTFGASCYTACP 300
Qy 301 YNLTSDVGSCTLVCPAHNQEVTAEDGTORCEKSKPCARVCYGLGMEHLREYRAVTSAN 360
Db 301 YNLTSDVGSCTLVCPAHNQEVTAEDGTORCEKSKPCARVCYGLGMEHLREYRAVTSAN 360
Qy 361 IOEFACCKRTFSGSLAFPLPSFDGDPASNTAPLQPEOLOVFELEITGTYLISAMPDSDLP 420
Db 361 IOEFACCKRTFSGSLAFPLPSFDGDPASNTAPLQPEOLOVFELEITGTYLISAMPDSDLP 420
Qy 421 DLSVFNOLQVIRGRIILHNAYSLTLOGIGISWIGLSLRELSGLALIHNNHLCFVHTV 480
Db 421 DLSVFNOLQVIRGRIILHNAYSLTLOGIGISWIGLSLRELSGLALIHNNHLCFVHTV 480
Qy 481 PMDOLFRNPHQALLHTANPEDECVGEGLACHQOLCARHGCMGPGPTQCVNCSQFLRGQEC 540
Db 481 PMDOLFRNPHQALLHTANPEDECVGEGLACHQOLCARHGCMGPGPTQCVNCSQFLRGQEC 540
Qy 541 VEECRVYGLCPREYVNAHCLPCHEPCOPONGSVTCFGEADQCVACAHYKDPPECVARC 600
Db 541 VEECRVYGLCPREYVNAHCLPCHEPCOPONGSVTCFGEADQCVACAHYKDPPECVARC 600
Qy 601 PSGVPRDLSTYMPIMKFPDEEGACQPCPINCTHSCVDLDDKGCAPDQASPLTSTISAVVG 660
Db 601 PSGVPRDLSTYMPIMKFPDEEGACQPCPINCTHSCVDLDDKGCAPDQASPLTSTISAVVG 660
Qy 654 ----- 653
Db 661 ILLVVLGVVFGILLIRROOKIRKRYTMRLLOETELVEPLTPSGAMPNOQMRILKETEL 720
Qy 654 ----- 653
Db 721 RKVKVLGSGAFGTYYKGIWIPDGENYKIPYAIKVLRENTSPKANKELIDEAYMAGVGP 780
Qy 654 ----- 653
Db 781 YVSRLLGICLTSTVOLVTOLMPYGCLLDHVRENRGLSGODLLNMCQIAKMSYLEDR 840
Qy 654 ----- 653

```

Db 841 LVHRDLAARNVLYKSPNHVKITDGLARLDDIDETEHADGKVPKMMALLESILRRRT 900  
QY 654 ----- 653  
Db 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPORPICTIDVYIMVCKMM 960  
QY 654 -----QNEDLGPASPLDSTFYRSLLDEDDMGDLVDA 684  
Db 961 IDSECRPRFRELYSEFSRMAKDPORFVVIQNEDLGPASPLDSTFYRSLLDEDDMGDLVDA 1020  
QY 685 EEYLVPOGFFCPCPPAPAGAGVHHRRSSSTRSGGDLTLGLPSEEARPSRLAPSEG 744  
Db 1021 EEYLVPOGFFCPCPPAPAGAGVHHRRSSSTRSGGDLTLGLPSEEARPSRLAPSEG 1080  
QY 745 AGSDVFDGDLGMGAKGLQSLPTHDPSLQRYSEDPVPLPSETDGVAPLTCSPQPEYV 804  
Db 1081 AGSDVFDGDLGMGAKGLQSLPTHDPSLQRYSEDPVPLPSETDGVAPLTCSPQPEYV 1140  
QY 805 NQPDVPRQPPSPREGPLPAARPAAGATLERPKTSLPGKNVYKDVFAAGAVENPEYLTPO 864  
Db 1141 NQPDVPRQPPSPREGPLPAARPAAGATLERPKTSLPGKNVYKDVFAAGAVENPEYLTPO 1200  
QY 865 GSAAPQHPAPSPADNLYYMDODPERGAPSTFKGPTAENPEYLGIDVPV 919  
Db 1201 GSAAPQHPAPSPADNLYYMDODPERGAPSTFKGPTAENPEYLGIDVPV 1255  
RES 6  
US-09-854-356-6  
Sequence 68: Application US/08468545B  
Patent No. 5876712  
GENERAL INFORMATION:  
APPLICANT: Cheever, Martin A.  
APPLICANT: Disis, Mary L.  
TITLE OF INVENTION: IMMUNE REACTIVITY TO HER-2/neu PROTEIN  
TITLE OF INVENTION: FOR DIAGNOSIS AND TREATMENT OF MALIGNANCIES IN WHICH THE  
TITLE OF INVENTION: HER-2/neu ONCOGENE IS ASSOCIATED  
NUMBER OF SEQUENCES: 69  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Seed and Berry LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
City: Seattle  
STATE: Washington  
COUNTRY: US  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/468,545B  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Sharkey, Richard G.  
REGISTRATION NUMBER: 32,629  
REFERENCE/DOCKET NUMBER: 920010.448C5  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 68:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1255 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-468-545B-68  
Query Match 96.3%; Score 4891; DB 2; Length 1255;  
Best Local Similarity 73.1%; Pred. No. 0;  
Matches 918; Conservative 0; Mismatches 1; Indels 336; Gaps 1;  
QY 1 MELALALCWGLLALLPPGAASVQVCTGTDKMLRLPASPETHDMLRLHLYQGGCVQVGNL 60

Db 1 MELALALCWGLLALLPPGAASVQVCTGTDKMLRLPASPETHDMLRLHLYQGGCVQVGNL 60  
QY 61 ELTYLPTNASLSEFQDIOEVQGVYLLAHNOVROYPLQRLIRVGTQLFEEDNYALAVLNDG 120  
Db 61 ELTYLPTNASLSEFQDIOEVQGVYLLAHNOVROYPLQRLIRVGTQLFEEDNYALAVLNDG 120  
QY 121 DPLNNTPTVYGASPGGIRELQSLTEILKGVYLQNNPQLCYODTLIMQDIFKNNQLA 180  
Db 121 DPLNNTPTVYGASPGGIRELQSLTEILKGVYLQNNPQLCYODTLIMQDIFKNNQLA 180  
QY 181 LTLIDTNRSRACHPCSPMGKSRGMSSEDDCSLTFTVCAGGCARKGPLPTDCHEOC 240  
Db 181 LTLIDTNRSRACHPCSPMGKSRGMSSEDDCSLTFTVCAGGCARKGPLPTDCHEOC 240  
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPLVYNTDTFESMPNDEGRYTFGASCVTACP 300  
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPLVYNTDTFESMPNDEGRYTFGASCVTACP 300  
QY 301 YNYLSTVGSCTLVCPILHNOEVTAEDETQRCCKSKFCARVYGLGMEHLREVAAYTSAN 360  
Db 301 YNYLSTVGSCTLVCPILHNOEVTAEDETQRCCKSKFCARVYGLGMEHLREVAAYTSAN 360  
QY 361 IOERAGCKKIFGSLAFIPESFDGDPASNTAPLOEQLOVEETLEITGYLISAMPDLP 420  
Db 361 IOERAGCKKIFGSLAFIPESFDGDPASNTAPLOEQLOVEETLEITGYLISAMPDLP 420  
QY 421 DLSYFQMLQVIRGIRILHNGAVSLTLOGISWLGRLSRLRELSGIALIHNHTLHCFVHTV 480  
Db 421 DLSYFQMLQVIRGIRILHNGAVSLTLOGISWLGRLSRLRELSGIALIHNHTLHCFVHTV 480  
QY 481 PMDOLFNRPHOALLHTANRPEDECVGSLACHOLACAGHCKGPPPTQCVNCSQFLRQEC 540  
Db 481 PMDOLFNRPHOALLHTANRPEDECVGSLACHOLACAGHCKGPPPTQCVNCSQFLRQEC 540  
QY 541 VEECRVLOGLPREYVNAHRCILCHPECOPONGSTGCPPEADQCVAAHYHDPFCVARC 600  
Db 541 VEECRVLOGLPREYVNAHRCILCHPECOPONGSTGCPPEADQCVAAHYHDPFCVARC 600  
QY 601 PSQVPRDLSTYPIWKFPDEBACQPCPINCTHSCVDLDDKCPAEOBAPLTS----- 653  
Db 601 PSQVPRDLSTYPIWKFPDEBACQPCPINCTHSCVDLDDKCPAEOBAPLTS----- 653  
QY 654 ----- 653  
Db 654 -----QNEDLGPASPLDSTFYRSLLDEDDMGDLVDA 684  
QY 685 EEYLVPOGFFCPCPPAPAGAGVHHRRSSSTRSGGDLTLGLPSEEARPSRLAPSEG 744  
Db 1021 EEYLVPOGFFCPCPPAPAGAGVHHRRSSSTRSGGDLTLGLPSEEARPSRLAPSEG 1080  
QY 745 AGSDVFDGDLGMGAKGLQSLPTHDPSLQRYSEDPVPLPSETDGVAPLTCSPQPEYV 804

Db 1081 AGSDVFDGLMGAKGLQSLPTHDPSLQRYSEDPVPLPSETDGYVAPLTCSPQPEYV 1140  
QY 805 NOPDVRPOPSPRPGPLPAARPAAGATLERPKTSLPGKNGYVKDVFAGGAVENPEYLTPQ 864  
Db 1141 NOPDVRPOPSPRPGPLPAARPAAGATLERPKTSLPGKNGYVKDVFAGGAVENPEYLTPQ 1200  
QY 865 GGAAPQHPHPPAPSPAFDNLTYMDODPPERGAPSPSTFKGTPTAENPEYLGLDVPV 919  
Db 1201 GGAAPQHPHPPAPSPAFDNLTYMDODPPERGAPSPSTFKGTPTAENPEYLGLDVPV 1255

RESULT 7  
US-08-466-680B-68  
; Sequence 68, Application US/08466680B  
; Patent No. 6075122  
; GENERAL INFORMATION:  
; APPLICANT: Cheever, Martin A.  
; APPLICANT: Disis, Mary L.  
; TITLE OF INVENTION: IMMUNE REACTIVITY TO HER-2/neu PROTEIN  
; TITLE OF INVENTION: FOR DIAGNOSIS AND TREATMENT OF MALIGNANCIES IN WHICH THE  
; TITLE OF INVENTION: HER-2/neu ONCOGENE IS ASSOCIATED  
; NUMBER OF SEQUENCES: 69  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Seed and Berry LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: US  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/466,680B  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sharkey, Richard G.  
; REGISTRATION NUMBER: 32,629  
; REFERENCE/DOCKET NUMBER: 920010.448C4  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 68:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1255 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
US-08-466-680B-68

Query Match 96.3%; Score 4891; DB 3; Length 1255;  
Best Local Similarity 73.1%; Pred. No. 0;  
Matches 918; Conservative 0; Mismatches 1; Indels 336; Gaps 1;

QY 1 MELAALCRWGLLALLPFGAASVYCTGTDMKRLPASPEYTHLMDLRLHYOGCOVYVGNL 60  
Db 1 MELAALCRWGLLALLPFGAASVYCTGTDMKRLPASPEYTHLMDLRLHYOGCOVYVGNL 60  
QY 61 ELTYLTPNASLFLQDIOEVGYVLIHNOVROYPLQRLRTVRGTOLFEENYALAVLDNG 120  
Db 61 ELTYLTPNASLFLQDIOEVGYVLIHNOVROYPLQRLRTVRGTOLFEENYALAVLDNG 120  
QY 121 DPLNNTTPVYGASPGGLRELQRLSLTEILKGVLIQIRNPOLCYODTILMKDIFKKNOLA 180  
Db 121 DPLNNTTPVYGASPGGLRELQRLSLTEILKGVLIQIRNPOLCYODTILMKDIFKKNOLA 180  
QY 181 LTLIDTNRSRACHPCSPMGKSGRCMGSSSEDCOSLRTVACGAGCARCKGPLPTDCCHEOC 240  
Db 181 LTLIDTNRSRACHPCSPMGKSGRCMGSSSEDCOSLRTVACGAGCARCKGPLPTDCCHEOC 240  
QY 241 AAGCTGPKHSDCLACLFHNHSGICELHCPALVTYNTDTFESMPNDEGRYTFGASCVYACP 300

Db 241 AAGCTGPKHSDCLACLFHNHSGICELHCPALVTYNTDTFESMPNDEGRYTFGASCVYACP 300  
QY 301 YNLTSDVGSCTIVCPHNDQVTAEDGTORCEKSKPCARCYGLGHEHLREVAVTYSAN 360  
Db 301 YNLTSDVGSCTIVCPHNDQVTAEDGTORCEKSKPCARCYGLGHEHLREVAVTYSAN 360  
QY 361 IOEFAGCKITFSLAFIPESFDGPASNTAPLOPEQLOVFETLEITGYLITSMPDLSL 420  
Db 361 IOEFAGCKITFSLAFIPESFDGPASNTAPLOPEQLOVFETLEITGYLITSMPDLSL 420  
QY 421 DLSVFQNLQVIRGRILNHGASLTLQGLIGISLWLSRLSRLSGLALIHNTHLFCFVHTV 480  
Db 421 DLSVFQNLQVIRGRILNHGASLTLQGLIGISLWLSRLSRLSGLALIHNTHLFCFVHTV 480  
QY 481 PMDQLFRNPQALLHTANRPDECEVGEGLACHQLCARGHCWPGPTQCVNCSQFLRGQEC 540  
Db 481 PMDQLFRNPQALLHTANRPDECEVGEGLACHQLCARGHCWPGPTQCVNCSQFLRGQEC 540  
QY 541 VEECRVILQGLPREVYNARHCLPCHECOPQNGSVTCGPEADQCVACAHKDPFCVYARC 600  
Db 541 VEECRVILQGLPREVYNARHCLPCHECOPQNGSVTCGPEADQCVACAHKDPFCVYARC 600  
QY 601 PSQVCPDLSTYMPIMKFPDEEGACOPCPINCTHSCVDLDDKCPAEOBASPLTS----- 653  
Db 601 PSQVCPDLSTYMPIMKFPDEEGACOPCPINCTHSCVDLDDKCPAEOBASPLTS----- 653  
QY 654 ----- 653  
Db 654 ----- 653  
QY 661 ILVVLGVVFGILIKRRQOKIRKRYTMRLLOETELVEPLTPSGAMPNQAOMRIKTEL 720  
Db 661 ILVVLGVVFGILIKRRQOKIRKRYTMRLLOETELVEPLTPSGAMPNQAOMRIKTEL 720  
QY 654 ----- 653  
Db 654 ----- 653  
QY 721 RKVKVLGSAFVYKGIWIPDGENVKI PAIKVIRENTSPKANKETIDEAYVAGVGP 780  
Db 721 RKVKVLGSAFVYKGIWIPDGENVKI PAIKVIRENTSPKANKETIDEAYVAGVGP 780  
QY 654 ----- 653  
Db 654 ----- 653  
QY 781 YVSRLLGICLSTVQVLTQMLPYGCLLDHVARENGRGLSQDLIMWCQIAKMSYLEDVR 840  
Db 781 YVSRLLGICLSTVQVLTQMLPYGCLLDHVARENGRGLSQDLIMWCQIAKMSYLEDVR 840  
QY 654 ----- 653  
Db 654 ----- 653  
QY 841 LVHRDLAARNVLYKSPNHVKITDFGLARLDDIDETEXHADGKVPKIMMALESLRRFT 900  
Db 841 LVHRDLAARNVLYKSPNHVKITDFGLARLDDIDETEXHADGKVPKIMMALESLRRFT 900  
QY 654 ----- 653  
Db 654 ----- 653  
QY 901 HQSDVMSYGVYWEIMTFGAKPYDGLPARETIDLEKGERLPOPICTIDVYIMVYKCM 960  
Db 901 HQSDVMSYGVYWEIMTFGAKPYDGLPARETIDLEKGERLPOPICTIDVYIMVYKCM 960  
QY 654 ----- 653  
Db 654 ----- 653  
QY 961 IDSECRPRFRELVSERMAARDPQRFVYIQNEDLGPASPLDSTFYKSLLEDDEDDGLVDA 1020  
Db 961 IDSECRPRFRELVSERMAARDPQRFVYIQNEDLGPASPLDSTFYKSLLEDDEDDGLVDA 1020  
QY 685 EBYLVPOGFECPDPAFAGAGVYHHRSSSTRSGGDLTLGLEPSEEAARSLAPSEG 744  
Db 1021 EBYLVPOGFECPDPAFAGAGVYHHRSSSTRSGGDLTLGLEPSEEAARSLAPSEG 744  
QY 745 AGSDVFDGLMGAKGLQSLPTHDPSLQRYSEDPVPLPSETDGYVAPLTCSPQPEYV 1140  
Db 1081 AGSDVFDGLMGAKGLQSLPTHDPSLQRYSEDPVPLPSETDGYVAPLTCSPQPEYV 1140  
QY 805 NOPDVRPOPSPRPGPLPAARPAAGATLERPKTSLPGKNGYVKDVFAGGAVENPEYLTPQ 864  
Db 1141 NOPDVRPOPSPRPGPLPAARPAAGATLERPKTSLPGKNGYVKDVFAGGAVENPEYLTPQ 1200  
QY 865 GGAAPQHPHPPAPSPAFDNLTYMDODPPERGAPSPSTFKGTPTAENPEYLGLDVPV 919  
Db 1201 GGAAPQHPHPPAPSPAFDNLTYMDODPPERGAPSPSTFKGTPTAENPEYLGLDVPV 1255

RESULT 8  
US-08-484-438-8  
; Sequence 8, Application US/08484438  
; Patent No. 5811098  
; Patent No. 5811098 5780031

GENERAL INFORMATION:  
APPLICANT: Plozman, Gregory D.  
APPLICANT: Culouscou, Jean-Michel  
APPLICANT: Shoyab, Mohammed  
APPLICANT: Stegall, Clay B.  
APPLICANT: Helistr m, Ingeerd  
APPLICANT: Helistr m, Karl E.  
TITLE OF INVENTION: HER4 HUMAN RECEPTOR TYROSINE KINASE  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/484,438  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/323,442  
FILING DATE: 14-OCT-1994  
APPLICATION NUMBER: US 08/150,704  
FILING DATE: 10-NOV-1993  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/981,165  
FILING DATE: 24-NOV-1992  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 5624-230  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741  
TELEX: 66141 PENNTE  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1255 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-484-438-8

Query Match 94.9%; Score 4818; DB 2; Length 1255;  
Best Local Similarity 72.6%; Pred No. 0;  
Matches 912; Conservative 0; Mismatches 6; Indels 338; Gaps 3;

QY 1 MELAALCRWGLLLALPPGAASYOCTGDMKRLRPASPEHLDMLRHLYOGCQVQGNL 60  
DB 1 MELAALCRWGLLLALPPGAASYOCTGDMKRLRPASPEHLDMLRHLYOGCQVQGNL 60  
QY 61 ELTYPTNASLSFDIOIQGVYLAHNOVQVPLQRLRIYRGTOLEFEDNVALAVLDNG 120  
DB 61 ELTYPTNASLSFDIOIQGVYLAHNOVQVPLQRLRIYRGTOLEFEDNVALAVLDNG 120  
QY 121 DELNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYODTILMKDIFHKNNOLA 180  
DB 121 DELNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYODTILMKDIFHKNNOLA 180  
QY 181 LTLIDTNRSRACHPCSPMKSGSRGSESESDCOSLTRVYVACAGACARCKGPLPTDCHEOC 240  
DB 181 LTLIDTNRSRACHPCSPMKSGSRGSESESDCOSLTRVYVACAGACARCKGPLPTDCHEOC 240  
QY 241 AAGCTGPKHSDCLACLFHNHSGICELHCPALVYNTDTFESMPNPGRYTFGASCVTACP 300

DB 241 AAGCTGPKHSDCLACLFHNHSGICELHCPALVYNTDTFESMPNPGRYTFGASCVTACP 300  
QY 301 YNTLSTDVSCITVYGLHNOEYAEQTORECECSPCARVCGGLGMEHLREVRAATSN 360  
DB 301 YNTLSTDVSCITVYGLHNOEYAEQTORECECSPCARVCGGLGMEHLREVRAATSN 360  
QY 361 IOEFACCKIFGSLAFPLPSFDGDPASNTAPLQPEOLQVFEETLEETGYLYISAMPDLP 420  
DB 361 IOEFACCKIFGSLAFPLPSFDGDPASNTAPLQPEOLQVFEETLEETGYLYISAMPDLP 420  
QY 421 DLSVFONLOYIRGRIILHNGAYSLLTLOGLSIWLGLSLRELSGGLLIHNNHLCVHTV 480  
DB 421 DLSVFONLOYIRGRIILHNGAYSLLTLOGLSIWLGLSLRELSGGLLIHNNHLCVHTV 480  
QY 481 PMDQLFRNPHOALHTANRDEDECVEGLACHOLCARGHMGVPTQCYNCSQFLRGOC 540  
DB 481 PMDQLFRNPHOALHTANRDEDECVEGLACHOLCARGHMGVPTQCYNCSQFLRGOC 540  
QY 541 VEECRVLOGLPREYVNAHRLCPHPECOPQNGSVTCFGEADQCVACAHYKDPPECVAC 600  
DB 541 VEECRVLOGLPREYVNAHRLCPHPECOPQNGSVTCFGEADQCVACAHYKDPPECVAC 600  
QY 601 PSGVKEPDLSTYMPYKPPDEGACQPCPINCTHSCVDLDDKGPABORASPLTS----- 653  
DB 601 PSGVKEPDLSTYMPYKPPDEGACQPCPINCTHSCVDLDDKGPABORASPLTSIYSAVYG 660  
QY 654 ----- 653  
DB 654 ILLVVVLGVVFGILIKRRQOKIRKTYMRLLQETELVEPLTPSGAMPNOQRILKETEL 720  
QY 654 ----- 653  
DB 721 RKVKVLGSGAFGVYKGIWIPGENYKIPAIKVLRENTSPANKREILDEAVYMAVGSP 780  
QY 654 ----- 653  
DB 781 YVSRLLIGLSTVOLVOTLMPYGLLDHVRENRGLGQDLLMCMQIAKGSYLEDV 840  
QY 654 ----- 653  
DB 841 LVHRDLAARNVLYKSPNNHYKITDFGLARLLIDETRYHADGGKVPITKMMALSLIRRET 900  
QY 654 ----- 653  
DB 901 HOSDWSYGVYVWELMTFGAKPYDGIPIAREIPDLLEKGRPLPQPICTIDVYIMVYKMM 960  
QY 654 ----- QNEIDGASPLDSTFFYSILLEDMDMDLVDA 684  
DB 961 IDSECRPRRELIVSEFSRMAKDPQRFVVTQNEIDLGPASPLDSTFFYSILLEDMDMDLVDA 1020  
QY 685 EEYLVPOQGFECDDPAPGAGWYHNRHSSSTRSGGDLTLGLPESEEPAPSPPLAPSEG 744  
DB 1021 EEYLVPOQGFECDDPAPGAGWYHNRHSSSTRSGGDLTLGLPESEEPAPSPPLAPSEG 1080  
QY 745 AGSDVEFDGLGMAKAGLOSLPTHDPSPLOKRYSEDPVPLPSETDGYVAPLTCSPQPEY 804  
DB 1081 AGSDVEFDGLGMAKAGLOSLPTHDPSPLOKRYSEDPVPLPSETDGYVAPLTCSPQPEY 1140  
QY 805 NOPDVAPQPPSPREGPLPAARAGATLERKTLSPKKNVYVDVAFGAVENREYLTQ 864  
DB 1141 NOPDVAPQPPSPREGPLPAARAGATLERKTLSPKKNVYVDVAFGAVENREYLTQ 1200  
QY 865 GGAAPQHPHPAFSPAFDNLVYWDOPPRGAPSPSTFKGPT-AENPEYIGLDVY 919  
DB 1201 GGAAPQHPHPAFSPAFDNLVYWDOPPRGAPSPSTFKGPTAENPEY-GLDVPY 1255

RESULT 9  
US-09-146-283-4  
; Sequence 4, Application US/09146283  
; Patent No. 5976546  
; GENERAL INFORMATION:

Query Match	71.5%	Score 3632	DB 2	Length 782
Query	Local Similarity	98.7%	Pred. No. 1.3e-271	Gaps
Matches	659	Conservative	0	Mismatches 5; Indels 4; Gaps 2
1	MELALICRMGLLALLLP	GAAS	TVCTGDMKLR	PASPETHLMDLRHLYGCGVYVGNL
1	MELALICRMGLLALLLP	GAAS	TVCTGDMKLR	PASPETHLMDLRHLYGCGVYVGNL
61	ELTYLPINASLFLDIO	EVGVYLLAH	NOVQVPLQRLIRIVR	GTQLEFDNTALAVLNG
61	ELTYLPINASLFLDIO	EVGVYLLAH	NOVQVPLQRLIRIVR	GTQLEFDNTALAVLNG
121	DLPLNTFTVTGASPC	GLKEIQLR	STELLKGVLLQOR	NQQLCYOPTIIMKDLFHKNNOLA
121	DLPLNTFTVTGASPC	GLKEIQLR	STELLKGVLLQOR	NQQLCYOPTIIMKDLFHKNNOLA
181	LTLIDITNRSRACHP	SPSPCKSGRCW	GESSEDDQS	LTRFYCAGACARCKGPLPTDCCHEQC
181	LTLIDITNRSRACHP	SPSPCKSGRCW	GESSEDDQS	LTRFYCAGACARCKGPLPTDCCHEQC
181	LTLIDITNRSRACHP	SPSPCKSGRCW	GESSEDDQS	LTRFYCAGACARCKGPLPTDCCHEQC
241	AAGCTGPKHSDCLAC	LHFHNSGICEL	ACPALVTVTDTFES	MNPDEGRYTGASCVTACP
241	AAGCTGPKHSDCLAC	LHFHNSGICEL	ACPALVTVTDTFES	MNPDEGRYTGASCVTACP
301	YNYLSTDVSGCTIVC	PLHNOEYTA	DGRORECKSK	PCARVCYGGIMEHLREVRAVTSAN
301	YNYLSTDVSGCTIVC	PLHNOEYTA	DGRORECKSK	PCARVCYGGIMEHLREVRAVTSAN
361	IOEFAGCKKIGSLA	FLPESFDG	PDASNTAP	LQAPOLQVETLEETIGLYISAPDSDLP
361	IOEFAGCKKIGSLA	FLPESFDG	PDASNTAP	LQAPOLQVETLEETIGLYISAPDSDLP
421	DLVSFQNLQVIRGR	ILHNGAVSL	TLQGLGISW	GLRSLRELGSGLALHHNTHLCFVHTV

Db	421	DLSTQNPQVIRGRIILNHGAATSLTLOGIGISWIGLRSLRELSSGSLAIHHNTHLCFPHY	480
Qy	481	PWDOLFRRPHALLHTANRPDEDECVGGSLACHQDLCANGHCWGRRPTQCVNCSQFLRGQEC	540
Db	481	PWDOLFRRPHALLHTANRPDEDECVGGSLACHQDLCANGHCWGRRPTQCVNCSQFLRGQEC	540
Qy	541	VEECGRVLOGLREXYVNNRHCILPCRHBCQOPONGSVTCGPEADOCVACAHYKDDPFVCAR	600
Db	541	VEECGRVLOGLREXYVNNRHCILPCRHBCQOPONGSVTCGPEADOCVACAHYKDDPFVCAR	600
Qy	601	PSGVKPDLSYMPIMKFPDEBGAQPCPTINCHSCVDLDDGGCAEORASPLTSONDILGP	660
Db	601	PSGVKPDLSYMPIMKFPDEBGAQPCPTINCHSCVDLDDGGCAEORASPLTSONDILGP	660
Qy	661	A-SPLDST 667	
Db	658	ARSPSPST 665	

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

QY 61 ELTYLPTNASSLSFLDIOIEVOGYVLLIAHNOVROPLORLRIYRGTOLEFEDNTALAVLDNG 120  
DB 61 ELTYLPTNASSLSFLDIOIEVOGYVLLIAHNOVROPLORLRIYRGTOLEFEDNTALAVLDNG 120  
QY 121 DPLNNTPTVTGASPGGLRELQRLSLTEILLKGVLIORNPOLCYODTILMKDIFHKNNOLA 180  
DB 121 DPLNNTPTVTGASPGGLRELQRLSLTEILLKGVLIORNPOLCYODTILMKDIFHKNNOLA 180  
QY 181 LFLIDNRSRACHPCSPMKSGSKMGESSEDDCSLRTYCAGGACARCKPPLPTDCHEOC 240  
DB 181 LFLIDNRSRACHPCSPMKSGSKMGESSEDDCSLRTYCAGGACARCKPPLPTDCHEOC 240  
QY 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDTESMPNPEGRTYFGASCYTACP 300  
DB 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDTESMPNPEGRTYFGASCYTACP 300  
QY 301 YNLTSDVGSCTLVCPHLHNOEYTAEDGTORCEKSPCARVCYGLGMEHLREYRAVTSAN 360  
DB 301 YNLTSDVGSCTLVCPHLHNOEYTAEDGTORCEKSPCARVCYGLGMEHLREYRAVTSAN 360  
QY 361 IOEFACCKRIFGSLAFLEPSFDGDPASNTAPLOPEOLOVFEETLEITGYLISAMPDSL 420  
DB 361 IOEFACCKRIFGSLAFLEPSFDGDPASNTAPLOPEOLOVFEETLEITGYLISAMPDSL 420  
QY 421 DLSVFONLOVIRGRILHNGAYSLTLOGLSIWLGLRSLRELSGSLALIHNNHLCFVHTV 480  
DB 421 DLSVFONLOVIRGRILHNGAYSLTLOGLSIWLGLRSLRELSGSLALIHNNHLCFVHTV 480  
QY 481 PMDOLFRNHQALHTANRPEDCEVGEGLACHQOLCARGHWGPGPQCVCNCSOFLGQEC 540  
DB 481 PMDOLFRNHQALHTANRPEDCEVGEGLACHQOLCARGHWGPGPQCVCNCSOFLGQEC 540  
QY 541 VEECRVLOGLPREYVNAHCLPCHPRECOPONGSVTCFGEADQCVAACHKDPFVCAR 600  
DB 541 VEECRVLOGLPREYVNAHCLPCHPRECOPONGSVTCFGEADQCVAACHKDPFVCAR 600  
QY 601 PSGVKDLSYMPIWKFPDEGACQPCPINCTHSCVDLDKGCAPAEORASPLTSQNEIDL 660  
DB 601 PSGVKDLSYMPIWKFPDEGACQPCPINCTHSCVDLDKGCAPAEORASPLTSLE--AP 657  
QY 661 A-SPLDST 667  
DB 658 ARSPSPST 665

RESULT 11  
US-09-344-195-4  
; Sequence 4, Application US/09344195  
; Patent No. 6210662  
; GENERAL INFORMATION:  
; APPLICANT: Laus, Reiner  
; Ruegg, Curtis L.  
; Mu, Hongyu  
; TITLE OF INVENTION: Immunostimulatory Compositions  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Dehlinger & Associates  
; STREET: 350 Cambridge Ave., Suite 250  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94306  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/344,195  
; FILING DATE: 24-Jun-1999  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/09/146,283

; FILING DATE: 03-SEPT-1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Judge, Linda R.  
; REGISTRATION NUMBER: 42,702  
; REFERENCE/DOCKET NUMBER: 7636-0010.21  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-324-0880  
; TELEFAX: 650-324-0960  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 782 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHEICAL: NO  
; ORIGINAL SOURCE:  
; ORGANISM: homo sapiens  
; INDIVIDUAL ISOLATE: GM-CSF-Her-2 fusion protein; Fig. 8  
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:  
US-09-344-195-4  
Query Match 71.5%; Score 3632; DB 4; Length 782;  
Best Local Similarity 98.7%; Pred. No. 1.3e-271;  
Matches 659; Conservative 0; Mismatches 5; Indels 4; Gaps 2;  
QY 1 MELALCRNGLLIALPPGAASTOVCTGDMKRLPASPETHIDMLRHLYOGQOVQGNL 60  
DB 1 MELALCRNGLLIALPPGAASTOVCTGDMKRLPASPETHIDMLRHLYOGQOVQGNL 60  
QY 61 ELTYLPTNASSLSFLDIOIEVOGYVLLIAHNOVROPLORLRIYRGTOLEFEDNTALAVLDNG 120  
DB 61 ELTYLPTNASSLSFLDIOIEVOGYVLLIAHNOVROPLORLRIYRGTOLEFEDNTALAVLDNG 120  
QY 121 DPLNNTPTVTGASPGGLRELQRLSLTEILLKGVLIORNPOLCYODTILMKDIFHKNNOLA 180  
DB 121 DPLNNTPTVTGASPGGLRELQRLSLTEILLKGVLIORNPOLCYODTILMKDIFHKNNOLA 180  
QY 181 LFLIDNRSRACHPCSPMKSGSKMGESSEDDCSLRTYCAGGACARCKPPLPTDCHEOC 240  
DB 181 LFLIDNRSRACHPCSPMKSGSKMGESSEDDCSLRTYCAGGACARCKPPLPTDCHEOC 240  
QY 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDTESMPNPEGRTYFGASCYTACP 300  
DB 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDTESMPNPEGRTYFGASCYTACP 300  
QY 301 YNLTSDVGSCTLVCPHLHNOEYTAEDGTORCEKSPCARVCYGLGMEHLREYRAVTSAN 360  
DB 301 YNLTSDVGSCTLVCPHLHNOEYTAEDGTORCEKSPCARVCYGLGMEHLREYRAVTSAN 360  
QY 361 IOEFACCKRIFGSLAFLEPSFDGDPASNTAPLOPEOLOVFEETLEITGYLISAMPDSL 420  
DB 361 IOEFACCKRIFGSLAFLEPSFDGDPASNTAPLOPEOLOVFEETLEITGYLISAMPDSL 420  
QY 421 DLSVFONLOVIRGRILHNGAYSLTLOGLSIWLGLRSLRELSGSLALIHNNHLCFVHTV 480  
DB 421 DLSVFONLOVIRGRILHNGAYSLTLOGLSIWLGLRSLRELSGSLALIHNNHLCFVHTV 480  
QY 481 PMDOLFRNHQALHTANRPEDCEVGEGLACHQOLCARGHWGPGPQCVCNCSOFLGQEC 540  
DB 481 PMDOLFRNHQALHTANRPEDCEVGEGLACHQOLCARGHWGPGPQCVCNCSOFLGQEC 540  
QY 541 VEECRVLOGLPREYVNAHCLPCHPRECOPONGSVTCFGEADQCVAACHKDPFVCAR 600  
DB 541 VEECRVLOGLPREYVNAHCLPCHPRECOPONGSVTCFGEADQCVAACHKDPFVCAR 600  
QY 601 PSGVKDLSYMPIWKFPDEGACQPCPINCTHSCVDLDKGCAPAEORASPLTSQNEIDL 660  
DB 601 PSGVKDLSYMPIWKFPDEGACQPCPINCTHSCVDLDKGCAPAEORASPLTSLE--AP 657  
QY 661 A-SPLDST 667  
DB 658 ARSPSPST 665



SEQUENCE CHARACTERISTICS:  
LENGTH: 624 amino acids  
TYPE: Amino Acid  
TOPOLOGY: Linear  
US-08-422-734-1

Query Match 68.4%; Score 3473; DB 4; Length 624;  
Best Local Similarity 99.8%; Pred. No. 1.7e-259;  
Matches 623; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
QY 22 STQVCTGDMKRLPASPETHLMDLRHLRYGCGVVGNNLETTYPTNASLFTDIDQEVQ 81
D 1 STQVCTGDMKRLPASPETHLMDLRHLRYGCGVVGNNLETTYPTNASLFTDIDQEVQ 60
QY 82 GYVLIANQVROYPLQRLRYRGTOLEFEDNYALAVLDNGDPRLNNTTPTVGSPPGLREIQ 141
D 61 GYVLIANQVROYPLQRLRYRGTOLEFEDNYALAVLDNGDPRLNNTTPTVGSPPGLREIQ 120
QY 142 LRSITELIKGGVLIQIRNPOLCYQDTILMKDIFHKNNQDLATLIDTNSRACHPCSPMCKG 201
D 121 LRSITELIKGGVLIQIRNPOLCYQDTILMKDIFHKNNQDLATLIDTNSRACHPCSPMCKG 180
QY 202 SRCGSESEDCQSLTRTVCAAGGACRCKGPLPTDCHEQCAAGCTGPKHSDLAQLHFNHS 261
D 181 SRCGSESEDCQSLTRTVCAAGGACRCKGPLPTDCHEQCAAGCTGPKHSDLAQLHFNHS 240
QY 262 GICELHCPALVTYNTDTFESMPNPEGRTFGASCVTACPYNYLSTDVGSCTLCPLHNOE 321
D 241 GICELHCPALVTYNTDTFESMPNPEGRTFGASCVTACPYNYLSTDVGSCTLCPLHNOE 300
QY 322 VTADGTQRECKSKPCARVCYGLGMEHLREVAVTSANIOEFAGCKKIFGSLAFLEPESF 381
D 301 VTADGTQRECKSKPCARVCYGLGMEHLREVAVTSANIOEFAGCKKIFGSLAFLEPESF 360
QY 382 DGDPASTAPLQPOLQVFELEETITGLYISAMPDLSPLDSVONQLOVINGRIILHNAY 441
D 361 DGDPASTAPLQPOLQVFELEETITGLYISAMPDLSPLDSVONQLOVINGRIILHNAY 420
QY 442 SLTLQGLISWLGRLSRLSGLALIHNTLHCFVHTVPMDQLFRNHQALLTANPE 501
D 421 SLTLQGLISWLGRLSRLSGLALIHNTLHCFVHTVPMDQLFRNHQALLTANPE 480
QY 502 DECVGEELACHQLCARHCHWPGPTQCVNCSQFLRGQECVECHVLOGLPREYVNAHCL 561
D 481 DECVGEELACHQLCARHCHWPGPTQCVNCSQFLRGQECVECHVLOGLPREYVNAHCL 540
QY 562 PCHEGCPONGSVTCFGEADQCAAHYKDPFCVACRPSGVKRPDLSYMPIMKFPDEEG 621
D 541 PCHEGCPONGSVTCFGEADQCAAHYKDPFCVACRPSGVKRPDLSYMPIMKFPDEEG 600
QY 622 ACQPCPINCTHSCVDLDDKGPAP 645
D 601 ACQPCPINCTHSCVDLDDKGPAP 624
```

RESULT 14  
US-09-630-155-2  
Sequence 2, Application US/09630155  
Patent No. 6414130

## GENERAL INFORMATION:

APPLICANT: Doherty, Joni Kristin and Gail M. Clinton  
TITLE OF INVENTION: HER-2 BINDING ANTAGONISTS  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: DAVIS WRIGHT TREMAINE LLP  
STREET: 1501 Fourth Avenue, 2600 Century Square  
CITY: Seattle  
STATE: Washington  
COUNTRY: U.S.A.  
ZIP: 98101  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: PC compatible

OPERATING SYSTEM: Windows95  
SOFTWARE: Word  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/630,155  
FILING DATE: 16-Jan-2001  
CLASSIFICATION: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Davison, Barry L.  
REGISTRATION NUMBER: 47,309  
REFERENCE/DOCKET NUMBER: 49321-10  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206 628-7621  
TELEFAX: 206 628-7699  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 419  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: unknown  
MOLECULE TYPE: polypeptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-09-630-155-2

Query Match 37.0%; Score 1878; DB 4; Length 419;  
Best Local Similarity 83.0%; Pred. No. 8.3e-137;  
Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;

```
QY 1 MELALCRWGLLLALPPGASTQVCTGDMKRLPASPETHLMDLRHLRYGCGVVGNNL 60
D 1 MELALCRWGLLLALPPGASTQVCTGDMKRLPASPETHLMDLRHLRYGCGVVGNNL 60
QY 61 ELTYLPTNASLFTDIDQEVQGYVLIANQVROYPLQRLRYRGTOLEFEDNYALAVLDNG 120
D 61 ELTYLPTNASLFTDIDQEVQGYVLIANQVROYPLQRLRYRGTOLEFEDNYALAVLDNG 120
QY 121 DPLNNTPTVYGASPGGLREQLRSITELIKGGVLIQIRNPOLCYQDTILMKDIFHKNNQDL 180
D 121 DPLNNTPTVYGASPGGLREQLRSITELIKGGVLIQIRNPOLCYQDTILMKDIFHKNNQDL 180
QY 181 LTLIDTNSRACHPCSPMCKSKRCMGSESEDCQSLTRTVCAAGGACRCKGPLPTDCHEQC 240
D 181 LTLIDTNSRACHPCSPMCKSKRCMGSESEDCQSLTRTVCAAGGACRCKGPLPTDCHEQC 240
QY 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRTFGASCVTACP 300
D 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRTFGASCVTACP 300
QY 301 YNYLSTDVGSCTLCPLHNOEVTADGTQRECKSKPCARVCYGLGMEHLREVAVTSAN 360
D 301 YNYLSTDVGSCTLCPLHNOEVTADGTQRECKSKPCARVCYGLGMEHLREVAVTSAN 360
QY 361 IOEFAGCKKIFGSLAFLEPESFGDPASTNT---APLOPOLQVFELEETITGLYISAMPD 417
D 361 IOEFAGCKKIFGSLAFLEPESFGDPASTNT---APLOPOLQVFELEETITGLYISAMPD 417
QY 418 SLPDLVSFQNLQVING 433
D 406 --PDAAHVAVNLNRYEG 419
```

RESULT 15  
US-08-484-438-7  
Sequence 7, Application US/0844438  
Patent No. 5811098

## GENERAL INFORMATION:

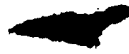
APPLICANT: PLOWMAN, Gregory D.  
APPLICANT: CULOUSCOU, Jean-Michel  
APPLICANT: SHOYAB, Mohammed  
APPLICANT: SIEGALL, Clay B.  
APPLICANT: HELISTR m, Ingegerd  
APPLICANT: HELISTR m, Karl E.  
TITLE OF INVENTION: HER4 HUMAN RECEPTOR TYROSINE KINASE



NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/484,438  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/323,442  
FILING DATE: 14-OCT-1994  
APPLICATION NUMBER: US 08/150,704  
FILING DATE: 10-NOV-1993  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/981,165  
FILING DATE: 24-NOV-1992  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 5624-230  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEO ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
\* LENGTH: 1210 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
U 8-484-438-7  
Query Match 32.4%; Score 1645; DB 2; Length 1210;  
Best Local Similarity 29.6%; P-adj. 118;  
Matches 374; Conservative 140; Mismatches 309; Indels 442; Gaps 22;

Db 363 TSISGDHLHPARSGSFHTTLPDQELDILKTVEITGFLIQQMPERDHLAFEN 422  
Qy 428 LQVIRGILHNGAVSLTLOGISWLGRLSLRELGSALIHNTILCFYHTVPMQDLFR 487  
Db 423 LEIRGRTOHQGFSLAVNSLINTSLGRSLKEISDDGVIIISGNKNCYANTIMKKLFG 482  
Qy 488 NPHQALLHTANRPDECVGSLACHQACARGHCHGPEPTQCVCNCSQFLRQECYEECRVL 547  
Db 483 TSGQKTKIISNGENSCATGQVCHALCSPEGCGPEPRCVCNRSRRCRECDKCKL 542  
Qy 548 QGUPREVYNAHCHPCRECPOMGVSVCGRPEADOCVACAHYKDPFCVAPRCSPKPD 607  
Db 543 EGPREFVENSSECIOCHPECLPQAMNITTCGRGDNKICQAHYIDGPHCVKTCPCAGVME 602  
Qy 608 LSTYPIWKFPDEEGACOPCTINCTHSCVDLDDKCPA----- 644  
Db 603 NNTL-VMKYADAGHVCHLCHPNTYCGTGGLCECPINGPKRIPSATGVALLLVVA 661  
Qy 645 -----EQRASPLTSQNE----- 656  
Db 662 LGIGLFMRHRHIVAKRTLRLRLDERLVEPLTPSGEAPNOALLRLKTEFEKKIKVLGSG 721  
Qy 657 ----- 656  
Db 722 AFGTVYKGLMIPBEKEKIVAIKELREATSPRANKRIIDEAYMASVDNPHVCRLLGIC 781  
Qy 657 ----- 656  
Db 782 LTSTVQLITQMPGCLLDVYREHKDNIQSQYLLNMCVQIAGKMYLEDRRLVHDLAAR 841  
Qy 657 -----DLG----- 659  
Db 842 NVLVKTPQHVKTDFGLAKLGAEEKEYHAEGKVPIKMALESILIRIYTHOSDVMYSY 901  
Qy 660 -----PA----- 661  
Db 902 VTYVELMTFSGSKPYDGIAPASEISILEKERLPQPPCTIDVYKIMYKCMHIDADSPPK 961  
Qy 662 -----SPLDSTFYRLSLEDDMDGLVDAEERYLPQ 692  
Db 962 RELIIEFSKMARDPORYLVIOGDERMHLPSPTDSNFYRALMEDMDVADADEYLIPQ 1021  
Qy 693 GFPPDPAPAGAGVHHHRHSSSTRSGGDLTLGLESESEEARSPRLAPSEGASDVFDG 752  
Db 1022 GFF-----SSPSTSTPLSSLSATSN--NS 1045  
Qy 753 DLGGAAGLQSLPTHDPSLQRYSEDPVLPSET--DGTVAPLTCSPQPEYVNPQDVR 810  
Db 1046 TVACIDRNGLOSCPIKEDSLQRYSSDPTGALTEDSIDDTFL-----PPEYINQ----- 1095  
Qy 811 PQQPPREGPLPARPAGATL-----ERPQLSPGKNGVYKDVAFGAVENPEYL--TPQ 864  
Db 1096 -----SVP--KRPAGSYQNPYYHNPPLNPAASRDPHYQD--PHSTAVANGPEYLTWQ 1143  
Qy 865 GGAAPQPHPPAPAPADNLVYMDQ-----DP-----PERGAPSTFKGTPTAE 908  
Db 1144 -----PTCVNSTSDPSAHMAQSGHOISLDNPDYQDPEPKKAPNGIKRS--TAE 1193  
Qy 909 NPEYL 913  
Db 1194 NAEYL 1198

Search completed: January 13, 2003, 14:49:51  
Job time : 28.7768 secs



•  
•  
•

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 13, 2003, 14:44:55 ; Search time 17.4617 Seconds  
(without alignments)  
3919.881 Million cell updates/sec

Title: US-09-854-356-7

Perfect score: 3954

Sequence: 1 MELDALCRMGLLALLPRGA.....GFCPPDPAPGAGVHHRRH 712

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: PIR.73:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3776	95.5	1255	1 A24571	protein-tyrosine k
2	3205	81.1	1260	1 TVRTMD	protein-tyrosine k
3	3185	80.6	1254	2 I48161	P-185 precursor -
4	1536	38.8	644	2 A36325	epidermal growth f
5	1534	38.8	1210	1 GQHUE	epidermal growth f
6	1533.5	38.8	1210	2 A53183	epidermal growth f
7	1509.5	38.2	1223	1 TVCHLV	epidermal growth f
8	1451	36.7	1308	2 A36223	epidermal growth f
9	1435	36.3	1342	2 A36223	kinase-related tra
10	1351.5	34.2	1339	2 JCA387	epidermal growth f
11	1313.5	33.2	1166	1 S06142	protein-tyrosine k
12	1155	29.2	527	2 A42032	epidermal growth f
13	1014.5	25.7	843	2 A27131	epidermal growth f
14	718.5	18.2	1323	2 E88237	protein let-23 [im
15	718.5	18.2	1374	2 S70712	protein-tyrosine k
16	691.5	17.5	1330	1 GOFPE	epidermal growth f
17	669.5	16.9	1369	2 S70713	protein-tyrosine k
18	621	15.7	1717	1 A45558	epidermal growth f
19	384	9.7	1363	2 T43220	insulin-like growt
20	370	9.4	1300	2 A36520	insulin receptor-t
21	370	9.4	1382	1 INHUR	insulin receptor p
22	363.5	9.2	366	2 D45558	epidermal growth f
23	362.5	9.2	1383	2 A36080	insulin receptor p
24	361	9.1	1372	2 A34157	insulin receptor p
25	353	8.9	1477	2 T18534	protein-tyrosine k
26	348.5	8.8	540	2 B47417	insulin receptor-t
27	338	8.5	1268	2 B36502	insulin receptor-t
28	331	8.4	333	2 B45558	epidermal growth f
29	331	8.4	342	2 C45558	epidermal growth f

30	326	8.2	1390	2 T30346	insulin receptor -
31	324	8.2	1607	2 T43212	insulin-like growt
32	316	8.0	1371	2 A33837	insulin-like growt
33	311.5	7.9	1367	1 IGHUR1	insulin-like growt
34	307	7.8	2101	2 S57245	insulin receptor (
35	307	7.8	2148	1 A56081	insulin receptor -
36	267.5	6.8	1846	2 T42047	insulin receptor h
37	244.5	6.2	1548	2 S34583	serine proteinase
38	240	6.1	1299	2 T43251	furin (EC 3.4.21.7
39	235	5.9	1329	2 A48805	insulin-like growt
40	226.5	5.7	183	2 JH0803	tyrosine kinase re
41	208	5.3	1680	2 A43434	furin (EC 3.4.21.7
42	195.5	4.9	937	2 I53282	gene PAC64 protein
43	195	4.9	932	2 I52527	PAC64A - mouse (fr
44	194.5	4.9	1574	2 T13954	MEGF6 protein - ra
45	191.5	4.8	915	1 A48225	subtilisin-like pr

#### ALIGNMENTS

RESULT 1  
A24571  
protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human  
N:Alternate names: c-erb-B-2 protein precursor; kinase-related transforming protein e  
C:Species: Homo sapiens (man)  
C:Date: 25-Oct-1987 #sequence\_revision 06-Dec-1996 #text\_change 11-Jun-1999  
C:Accession: A24571; A25491; A44188; B44188; I59509; I57622  
R:Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; Salto, T  
Nature 319, 230-234, 1986  
A:Title: Similarity of protein encoded by the human c-erb-B-2 gene to epidermal growt  
A:Reference number: A24571; MUID:86118663; PMID:3003577  
A:Accession: A24571  
A:Molecule type: mRNA  
A:Residues: 1-1255 <YAM>  
A:Cross-references: GB:X03363; NID:g31197; PIDN:CAA27060.1; PID:g31198  
R:Semba, K.; Kamata, N.; Toyoshima, K.; Yamamoto, T.  
Proc. Natl. Acad. Sci. U.S.A. 82, 6497-6501, 1985  
A:Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/epid  
A:Reference number: A25491; MUID:86016729; PMID:2995967  
A:Accession: A25491  
A:Molecule type: DNA  
A:Residues: 737-1031 <SEM>  
A:Cross-references: GB:M11767; NID:g182163; PIDN:AAA35808.1; PID:g553282  
R:Coussens, L.; Yang-Feng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seeburg  
Science 230, 1132-1139, 1985  
A:Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares chro  
A:Reference number: A44188; MUID:86070181; PMID:2999974  
A:Accession: A44188  
A:Molecule type: DNA  
A:Residues: 740-910 <COU1>  
A:Cross-references: GB:M12036; NID:g183988; PIDN:AAA35978.1; PID:g183989  
A:Accession: B44188  
A:Molecule type: mRNA  
A:Residues: 1-517, 'RAL', 522, 'S', 524-654, 'V', 656-1169, 'A', 1171-1255 <COU2>  
A:Cross-references: GB:M11750; NID:g183986  
R:King, C.R.; Kraus, M.H.; Aaronson, S.A.  
Science 229, 974-976, 1985  
A:Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma.  
A:Reference number: I59509; MUID:85272597; PMID:2992089  
A:Accession: I59509  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 832-909 <REX>  
A:Cross-references: GB:L29395; NID:g459807; PIDN:AAA35809.1; PID:g459808  
R:Tal, M.; King, C.R.; Kraus, M.H.; Ulrich, A.; Schlesinger, J.; Givol, D.  
Mol. Cell. Biol. 7, 2597-2601, 1987  
A:Title: Human HER2 (neu) promoter: evidence for multiple mechanisms for transcriptio  
A:Reference number: I57622; MUID:87286898; PMID:3039351  
A:Accession: I57622  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-191 <TAI>

A:Cross-references: GB:M16792; NID:9183983; PIDN:AAA58637.1; PID:9553332  
 C:Comment: Amplification and overexpression of this erbB-related gene occurs in about 30  
 C:Genetics:  
 A:Gene: GDB:ERBB2; NGL; NEU; HER-2  
 A:Cross-references: GDB:120613; OMIM:164870  
 A:Map position: 17q21.1-17q21.1  
 A:Introns: 25/1; 75/3; 147/1; 883/3  
 A>Note: the list of introns is incomplete  
 C:Function:  
 A:Description: catalyzes the phosphorylation of a peptidyl tyrosine residue by ATP  
 C:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho  
 inase  
 F:1-21/Domain: signal sequence #status predicted <SIG>  
 F:22-1255/Product: protein-tyrosine kinase erbB2 #status predicted <MAT>  
 F:22-653/Domain: extracellular #status predicted <EXT>  
 F:70-304/Domain: EGF receptor extracellular domain repeat <EE1>  
 F:395-605/Domain: EGF receptor extracellular domain repeat <EE2>  
 F:654-675/Domain: transmembrane #status predicted <TM>  
 F:676-1255/Domain: intracellular #status predicted <INT>  
 F:718-983/Domain: protein kinase homology <KIN>  
 F:726-734/Region: protein kinase ATP-binding motif  
 F:68,124,187,259,530,571,629/Binding site: carbohydrate (Asn) (covalent) #status predict  
 F:686/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted  
 F:753/Active site: Lys #status predicted  
 F:1139,1221,1222,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

Query Match Score 95.5%; Score 3776; DB 1; Length 1255;  
 Best Local Similarity 67.9%; Pred. No. 3.8e-244;  
 Matches 712; Conservative 0; Mismatches 0; Indels 336; Gaps 1;

```

QY 1 MELAALCRMGILLALPPGAASSTOVCTGTMKRLPASPEHLMHLYOGCQVVGNTL
Db 1 MELAALCRMGILLALPPGAASSTOVCTGTMKRLPASPEHLMHLYOGCQVVGNTL 60
QY 61 ELTYLPNNAISLFQDIQEVGYVLAHNOYRQVPLQRLKIVKSTQLFEDNYALAVDNG 120
Db 61 ELTYLPNNAISLFQDIQEVGYVLAHNOYRQVPLQRLKIVKSTQLFEDNYALAVDNG 120
QY 121 DPLNNTPTVPGASPGGRLREQLSLTEILKGVLIQNPOLCYODTILMDIFKNNOLA 180
Db 121 DPLNNTPTVPGASPGGRLREQLSLTEILKGVLIQNPOLCYODTILMDIFKNNOLA 180
QY 121 DPLNNTPTVPGASPGGRLREQLSLTEILKGVLIQNPOLCYODTILMDIFKNNOLA 180
Db 121 DPLNNTPTVPGASPGGRLREQLSLTEILKGVLIQNPOLCYODTILMDIFKNNOLA 180
QY 181 LTLIDNRSRACHPCSPMCKSGKSSSDCSLTFTVAGGACARCKGKPLPTDCHEQC 240
Db 181 LTLIDNRSRACHPCSPMCKSGKSSSDCSLTFTVAGGACARCKGKPLPTDCHEQC 240
QY 181 LTLIDNRSRACHPCSPMCKSGKSSSDCSLTFTVAGGACARCKGKPLPTDCHEQC 240
Db 181 LTLIDNRSRACHPCSPMCKSGKSSSDCSLTFTVAGGACARCKGKPLPTDCHEQC 240
QY 241 AAGCTGKSHDCLACLFHNSGICELHCPALVYNTDTFESMPNDEGRYTFGASCYACP 300
Db 241 AAGCTGKSHDCLACLFHNSGICELHCPALVYNTDTFESMPNDEGRYTFGASCYACP 300
QY 241 AAGCTGKSHDCLACLFHNSGICELHCPALVYNTDTFESMPNDEGRYTFGASCYACP 300
Db 241 AAGCTGKSHDCLACLFHNSGICELHCPALVYNTDTFESMPNDEGRYTFGASCYACP 300
QY 301 YNTLSTVSGCTVCPPLHNOEVAEDSTORCEKSKPCARCVGLGMEHLREVAVVSAN 360
Db 301 YNTLSTVSGCTVCPPLHNOEVAEDSTORCEKSKPCARCVGLGMEHLREVAVVSAN 360
QY 301 YNTLSTVSGCTVCPPLHNOEVAEDSTORCEKSKPCARCVGLGMEHLREVAVVSAN 360
Db 301 YNTLSTVSGCTVCPPLHNOEVAEDSTORCEKSKPCARCVGLGMEHLREVAVVSAN 360
QY 361 IOEFAGCKITFGSLAFPESEFGDPASNTAPLQPEQLQVETTELITGYLYISAMPDLP 420
Db 361 IOEFAGCKITFGSLAFPESEFGDPASNTAPLQPEQLQVETTELITGYLYISAMPDLP 420
QY 361 IOEFAGCKITFGSLAFPESEFGDPASNTAPLQPEQLQVETTELITGYLYISAMPDLP 420
Db 361 IOEFAGCKITFGSLAFPESEFGDPASNTAPLQPEQLQVETTELITGYLYISAMPDLP 420
QY 421 DLSYFQMLQYIRGILNNGAYSLTQGLGISMGLRSLRELGSGIALIHNTHTLCFVHTV 480
Db 421 DLSYFQMLQYIRGILNNGAYSLTQGLGISMGLRSLRELGSGIALIHNTHTLCFVHTV 480
QY 421 DLSYFQMLQYIRGILNNGAYSLTQGLGISMGLRSLRELGSGIALIHNTHTLCFVHTV 480
Db 421 DLSYFQMLQYIRGILNNGAYSLTQGLGISMGLRSLRELGSGIALIHNTHTLCFVHTV 480
QY 481 PMDOLFENPHOALHTNRPDEDECVGEGILACHOLCARGHCHGPPPTOCVNCISOFLRQDEC 540
Db 481 PMDOLFENPHOALHTNRPDEDECVGEGILACHOLCARGHCHGPPPTOCVNCISOFLRQDEC 540
QY 481 PMDOLFENPHOALHTNRPDEDECVGEGILACHOLCARGHCHGPPPTOCVNCISOFLRQDEC 540
Db 481 PMDOLFENPHOALHTNRPDEDECVGEGILACHOLCARGHCHGPPPTOCVNCISOFLRQDEC 540
QY 541 VEECRVQLGLPREYVNAHCLPCHPECOPOGNGSYTCFPEADQCVACAHYKDPFCVARC 600
Db 541 VEECRVQLGLPREYVNAHCLPCHPECOPOGNGSYTCFPEADQCVACAHYKDPFCVARC 600
QY 541 VEECRVQLGLPREYVNAHCLPCHPECOPOGNGSYTCFPEADQCVACAHYKDPFCVARC 600
Db 541 VEECRVQLGLPREYVNAHCLPCHPECOPOGNGSYTCFPEADQCVACAHYKDPFCVARC 600
QY 601 PSGVKPDLSTYPMKPPDEEGACQPCPINCTHSCVDLDDKCPAQRASPLTSSAAYVG 660
Db 601 PSGVKPDLSTYPMKPPDEEGACQPCPINCTHSCVDLDDKCPAQRASPLTSSAAYVG 660
QY 654 ----- 653
Db 661 ILLVVLGVVFGILLIKRROOKIKKTYMRLLQETELVEPLTPSGAMPNQAQRILKETEL 720
QY 654 ----- 653
Db 721 RKVKVLGSGAFGVYKGIWIPQGENVKIPALIKLRNENTSPKANKETILDEAYVMAGVSP 780
QY 654 ----- 653
Db 781 YSRLLGICLTSTVOLVTOIMPYCCLLDHYENRGRLSQDLLMCMQIAGKMSYLEDVR 840
QY 654 ----- 653
Db 841 LVHRLAARNVLVKSPNHVKITDPGLARLLDIDETEHADGKVPIKMALESILRRRT 900
QY 654 ----- 653
Db 901 HQSDVSYGVTVWELMTFGAKPYDIPAREIPDLLEKGERLPQPICTIDVYIMVYKCMW 960
QY 654 ----- 684
Db 961 IDECRPRFRELYSEFSRMAADPQREVVIONEDLPASPLDSTFYSLLLEDMDGLVDA 1020
QY 685 EEYLVPOQGFPCPDPAAGAGVHHRR 712
Db 1021 EEYLVPOQGFPCPDPAAGAGVHHRR 1048

```

Db 601 PSGVKPDLSTYPMKPPDEEGACQPCPINCTHSCVDLDDKCPAQRASPLTSSAAYVG 660  
 QY 654 ----- 653  
 Db 661 ILLVVLGVVFGILLIKRROOKIKKTYMRLLQETELVEPLTPSGAMPNQAQRILKETEL 720  
 QY 654 ----- 653  
 Db 721 RKVKVLGSGAFGVYKGIWIPQGENVKIPALIKLRNENTSPKANKETILDEAYVMAGVSP 780  
 QY 654 ----- 653  
 Db 781 YSRLLGICLTSTVOLVTOIMPYCCLLDHYENRGRLSQDLLMCMQIAGKMSYLEDVR 840  
 QY 654 ----- 653  
 Db 841 LVHRLAARNVLVKSPNHVKITDPGLARLLDIDETEHADGKVPIKMALESILRRRT 900  
 QY 654 ----- 653  
 Db 901 HQSDVSYGVTVWELMTFGAKPYDIPAREIPDLLEKGERLPQPICTIDVYIMVYKCMW 960  
 QY 654 ----- 684  
 Db 961 IDECRPRFRELYSEFSRMAADPQREVVIONEDLPASPLDSTFYSLLLEDMDGLVDA 1020  
 QY 685 EEYLVPOQGFPCPDPAAGAGVHHRR 712  
 Db 1021 EEYLVPOQGFPCPDPAAGAGVHHRR 1048

## RESULT 2

## TVRTNU

protein-tyrosine kinase (EC 2.7.1.112) neu precursor - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 31-Dec-1988 #sequence\_revision 31-Dec-1988 #text\_change 11-Jun-1999

C:Accession: A24562; A61204

R:Bargmann, C.I.; Hung, M.C.; Weinberg, R.A.

Nature 319, 226-230, 1986

A:Title: The neu oncogene encodes an epidermal growth factor receptor-related protein

A:Reference number: A24562; MUID:86118662; PMID:3945311

A:Accession: A24562

A:Molecule type: mRNA

A:Residues: 1-1260 <BAR>

A:Cross-references: EMBL:X03362; NID:956745; PIDN:CAA27059.1; PID:956746

R:Mesu, T.; Mann, A.M.; Macatee, T.L.; Garland, E.M.; Okamura, T.; Smith, R.A.; Cohe

Cartiogenesis 12, 1975-1978, 1991

A:Title: Direct DNA sequencing of the rat neu oncogene transmembrane domain reveals n

2-thiazolyl]formamide or N-methyl-N-nitrosourea.

A:Reference number: A61204; MUID:92035293; PMID:1682063

A:Accession: A61204

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 637-663, 'V', 665-702 <MAS>

A>Note: authors translated the codon GCA for residue 25 as Val

C:Genetics:

A:Gene: neu

C:Superfamily: epidermal growth factor receptor; protein kinase homology

C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phos

F:1-19/Domain: signal sequence #status predicted <SIG>

F:20-1260/Product: protein-tyrosine kinase neu #status predicted <MAT>

F:658-680/Domain: transmembrane #status predicted <TM>

F:723-988/Domain: protein kinase homology <KIN>

F:731-739/Region: protein kinase ATP-binding motif

F:71,191,263,535,576,634/Binding site: carbohydrate (Asn) (covalent) #status predicted

F:661/Binding site: phosphate (Thr) (covalent) #status predicted

F:758/Active site: Lys #status predicted

F:1227,1253/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match Score 3205; DB 1; Length 1260;  
 Best Local Similarity 57.9%; Pred. No. 5e-206;  
 Matches 608; Conservative 34; Mismatches 70; Indels 338; Gaps 3;

QY	1	MEALICRMGLLALLAPRGASNPQVCGTDMKRLPASEPHLDMIRHLXOCCOVVGNL	60
Db	4	MELAMCRMGWGLLALLPRGAGTAGTQVCTGDMKRLPASEPHLDMIRHLXOCCOVVGNL	63
QY	61	ELTYLPTNASSLSEFJODIQEOGVYLLAHNOVROPRLQRLRYRGQLEFDNALAVLNG	120
Db	64	ELTYVPANASSLSEFJODIQEOGVYLLAHNOVKRVLQRLRYRGQLEFDKALAVLNR	123
QY	121	DPLNNTTPVT-GASPGGLRELQRLSLTELLKGGVLIQNRPLCYQDTLMDIEHKNQL	179
Db	124	DPOONVAASPRGRTPEGLRELQRLSLTELLKGGVLIQNRPLCYQDMVLMKDVFRKNQL	183
QY	180	ALTDINBRACHPCSPMKGSKRGCGESEDQSLTRVCAGGACRGRLPTCCHEQ	239
Db	184	APVDIDINBRACHPCSPACKMDHMGESBEDQLTGTITSGCARCKGRPLTCCHEQ	243
QY	240	CAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDFEEMPNPEGRHYTGASCVCAC	299
Db	244	CAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDFEEMHNDEGRHYTGASCVCAC	303
QY	300	PYNTLSTDVQSSCLNCPRLAHQEBTAEQGRQKCKSPCARVCYGLGIMENLREVAUTISA	359
Db	304	PYNTLSTEVQSSCLNCPRNQEBTAEQGRQKCKSPCARVCYGLGIMENLREVAUTISD	363
QY	360	NIOFACCKRIEFSLAFLEPSEFQDASMTAPLQPOLQVFEETLEITGYLISAMPSTL	419
Db	364	NIOFEOCKRIEFSLAFLEPSEFQDSSGIAPLRPQLOVFEETLEITGYLISAMPSTL	423
QY	420	PDLVSFQNLQVINGRILLHNGAVSLTLQGLISWLGSLRSLRELGSGLALIHNTHLCPVHT	479
Db	424	RDLVSFQNLRIINGRILLHNGAVSLTLQGLISHLGSLRSLRELGSGLALIHRAHLCPVHT	483
QY	480	VPMQDLFRNHQALLHTANRDE-CVGGSLACHOLCAGHGWGSGPQCVVCSQFLNGQ	538
Db	484	VPMQDLFRNHQALLHSGNPEEDLCVSSGLVCNSLCAHGWGSGPQCVVCSHFLLGQ	543
QY	539	ECVEECFVLOGLEPREYVNAHCLPCHPEQOPQNGSVTCHGEPAQOCVACAHYKDRPFCVA	598
Db	544	ECVEECFVWMLGPREYVSDKRLCPHPEQOPQNSBTCHGSADQCAACAHYKDDSSCVA	603
QY	599	RCBSGVCPDLSTYMPIMKFPDEBQACOPCPTINCTHSCVDLDDKGCRAQDASPLT-----	652
Db	604	RCBSGVCPDLSTYMPIMKYPRDEBQICQPCPTINCTHSCVDLDDKGCRAQDASPLTIIATV	663
QY	653	-----	652
Db	664	EGVLLFLILVVVGLIKRRQKIRKTYMKRLLDQETELVEPLTPBGAHPNOAMRIKET	723
QY	653	-----	652
Db	724	ELARKVKLGSAGAGFYKGIWIPDGENVKIPVAIKVLRENTSPPANKETLDEAYVMAGV	783
QY	653	-----	652
Db	784	SPYVRLGLIGLSTVOLYTQIMPYCCLDHWREHGRGLSODLLMWCQIAKGSYLED	843
QY	653	-----	652
Db	844	VRLVHDLAARNVLVKSPIHVKLTDEGLARLDDIBETEVHADGKAVPLIKMMALJSLRR	903
QY	653	-----	652
Db	904	FTHOSDWSYGVTVWELMTFGAKPYXDIGIPARETLDLLEKGERLPPRPICTIDVYVIMKC	963
QY	653	-----SONEDLGASPLDSTFYRSLLEDDMGVLY	682
Db	964	WMIDSECRPRFRELVESEFMAKNDPRFVYLIQNEIDGSPSPDSTFYRSLLEDDMGVLY	1023
QY	683	DAEETLVPOQGFCCPRAPGAGAGMHNR	712
Db	1024	DAEETLVPOQGFSSPDTPTGTGTAAHRRR	1053

Result 3  
148161  
P:185 precursor - golden hamster  
C:Species: Mesocricetus auratus (golden hamster)  
C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 18-Jun-1999  
C:Accession: 148161  
R:Nakamura, T.; Ushijima, T.; Ishizaka, Y.; Nageo, M.; Arai, M.; Yamazaki, Y.; Ishike  
Gene 140, 251-255, 1994  
A>Title: Cloning and activation of the Syrian hamster neu proto-oncogene.  
A:Reference number: 148161; MUID:94193007; PMID:7908275  
A:Accession: 148161  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1254 <RES>  
A:Cross-references: GB:D16295; NID:g493236; PIDN:BAA03801.1; PID:g747595  
C:Genetics:  
A:Gene: neu  
C:Superfamily: epidermal growth factor receptor; protein kinase homology  
C:Keywords: ATP  
F:718-983/Domain: protein kinase homology <KIN>  
E:726-734/Region: protein kinase ATP-binding motif

Query Match	80.6%	Score 3185	DB 2	Length 1254
Best Local Similarity	57.5%	Pred. No. 1,1e-204		
Matches 603	Conservative	36	Indels 23	Gaps 336
1	MELAAALCRGGLLALLALPPAASPOVCTGDMKLRLPASPETHLMDLRHLXOGCOVYQGM	60		
1	MELAAACMGGLLALLSPGASGTQCTGDMKLRLPASPETHLMDLRHLXOGCOVYQGM	60		
61	ELTYLPNALSFLADIOEVOGVYLIANNOVRLPQRLRLRYGTOLFEDNYALAVLDNG	120		
61	ELTYLPANATLSPLADIOEVOGYMLIAHSQVHRVLPQRLRLRYGTOLFEDKAYALAVLDR	120		
121	DEPLNTPYTGASPGGLRELOLRSLTEILKGVLIQVLIQNRNQLCYODTPIIMKDFHKNQOLA	180		
121	DEPLNTPYTGATPEGLRELOLRSLTEILKGVLIQVLIQNRNQLCYODTPIIMKDFHKNQOLA	180		
181	LVLIDPNRSRAOHPSPMCKGSRMGESSEDDOSLTRPVAGAGCARCKRPLTDDCHEOC	240		
181	PVDIDTNSRACRPPCARPACKNDHMGASDEDDOITLTGITAPAVPAARARLPLTDDCHEOC	240		
241	AAGCTGPKHSDCIACIHFNHSGICELHCPALVTYNTDFESMNPDEGRATFGASCVTACR	300		
241	AAGCTGPKHSDCIACIHFNHSGICELHCPALVTYNTDFESMNPDEGRATFGASCVTACR	300		
301	YVLTSDNGSCILVCPRLANOEVYTAEDGTORCEKSKPCARVYGTGEMELREYRAVTSAN	360		
301	YVLTSTEVGSCILVCPRLANOEVYTAEDGTORCEKSKPCARVYGTGEMELREYRAVTSAN	360		
361	IOEPFAGCKITFESLAFPLPSPFDGDASNTARLPORPOLQVFEETLEETGLYISAMPDSLAP	420		
361	IOEPFAGCKITFESLAFPLPSPFDGDASNTARLPORPOLQVFEETLEETGLYISAMPDSLAP	420		
421	DLVSFQNLQVINGRILAHNGAVSLTLOGLGISWLGSRSLSELSGSLAIHHNTHLCFVHTY	480		
421	DLVSFQNLQVINGRILAHNGAVSLTLOGLGISWLGSRSLSELSGSLAIHHNTHLCFVHTY	480		
481	PMDOJLFRNHQALHTANRPDEECYGBGLACIQCACRGHCWGPRPTOCVCSQTLRGQBC	540		
481	PMDOJLFRNHQALHTANRPDEECYGBGLACIQCACRGHCWGPRPTOCVCSQTLRGQBC	540		
541	VEECVAVLQGLPREYANARICLPCRPBCOPONSVMCFGEADQCAVACAHYKORPPCVARC	600		
541	VEECVAVLQGLPREYANARICLPCRPBCOPONSVMCFGEADQCAVACAHYKORPPCVARC	600		
601	PSGVAPDLSYMPIMEFPDEGACQCPINCTHSCVDLDDKGCAPABORASPLNS	660		
601	PSGVAPDLSYMPIMEFPDEGACQCPINCTHSCVDLDDKGCAPABORASPLNS	660		
654		653		
661	ILFLVIGVVGILIKRRQKIRKTYMRRLQETELVEPLPSSGAMPQAOQMRILKETEL	720		

```

QY 654 ----- 653
Db 721 RKAVVLGSGAGFYKKGWIPDGENVKIPVAIKVLRNENTSKANKELLDEAYVWAGLSP 780
QY 654 ----- 653
Db 781 YSRHLGICLITVQVLTOLMPYGLLDHVEHRRGLGSODLLMWQVQIAKMSYLEDVDR 840
QY 654 ----- 653
Db 841 LVHMDLAARNVLYKSPNHVKTITDFGLARLLDIDETEHADGKVPDKMIALESLIRRET 900
QY 654 ----- 653
Db 901 HQSDWSYGYTWELMTFGAKPYDIPARELPDLEKGERLPQPPICIDVYIMVNCWM 960
QY 654 ----- 684
Db 961 IDSECRPRFRELVSERFARMARDPQRFVYVIONEDLGSSPLDSTFYSRLLEDMDGLVDA 1020
QY 685 EEVLVPOGFEPCPDAPGAGMHHRR 712
Db 1021 EEVLVPOGFEPCPDAPGAGMHHRR 1048

RESULT 4
A36325
epidermal growth factor receptor - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 25-Jan-1991 #sequence, revision 25-Jan-1991 #text, change 10-Oct-1997
C:Accession: A36325
R:Patch, L.A.; Harris, J.; Raymond, V.W.; Blasband, A.; Lee, D.C.; Earp, H.S.
Mol. Cell. Biol. 10, 2973-2982, 1990
A:Title: A truncated, secreted form of the epidermal growth factor receptor is encoded b
A:Reference number: A36325; MUID:90258886; PMID:2342466
A:Accession: A36325
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-644 <P>
A:Cross-references: GB:M37394
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: alternative splicing; ATP; growth factor receptor

Query Match 38.8%; Score 1536; DB 2; Length 644;
Best Local Similarity 44.6%; Pident. No. 66-95;
Matches 287; Conservative 96; Mismatches 234; Indels 24; Gaps 8;

QY 3 LAALCRWGLLALLPGA-ASTQVCTGTDMKRLPASPETHLDMRLHYOGCQVQCNLE 61
Db 15 LAALCAAG-----GALKEKVKCGTNSRLTQLGTFEDHFLSLQRMFNCCVYVNLLE 66
QY 62 LTYPTNASTFLDIDIOVGYVLIHNOVAQVPLQRIKRYGQLEPNDNALAVLNDGD 121
Db 67 ITTYVRYVDSFLKTIQEVAGYVLIATNTERIPLENIQIIRGNALYENYALVLSN-- 124
QY 122 PLNNTTPTVGTASPGRELOLRSLEIKGGVLIQORNPOLYQDPIIMKODIFHNOLAL 181
Db 125 -----YGNKTGLKELPRKNIQELIIGAVRSNNPILICNMETIQRDIV-ODVFLSN 175
QY 182 TLIDTNRN-RACHPCSPMKSGKRCMGSSSDQSLRTVTCAGGCA-RCKGPPLTDDCHEQ 239
Db 176 MSMDVQHLNGCPCDPSRCMGSGMBEHCQKLTIKICQAGSRRGRGSPDDCCINQ 235
QY 240 CAAGCTPRKISDCLACIHPNHSICELHCPALVYNTDFESMPNPGRYTGGASCTAC 299
Db 236 CAAGCTPRPSDCLYCHRFREDEATCKDTCPLMLYNPTTYQMDVNPBGKYSFGATCYVKK 295
QY 300 PYNTLSIDVSGSLVCLHNOEYTAEDGTQRCCKSPCAVYCGLGMCHLRVRAVYSA 359
Db 296 PRNVVYVDHSGSVACAPDPYIEV-EEGVSCKCKKCDQCRKRVGNGIGERKDLISNAT 354
QY 360 NIDFACCKIKFGSLAFPSFDGDPASNTAPLQPEQVQVETLEITGYLYISAMPDSL 419

```

```

Db 355 NIKHFKYCTAISGDHLHPVAFKGDSEFTRRPDLDPRELEILKTVKEITGFLIQAMPENW 414
QY 420 PDLSPFQNLQYIRGRILHNGAVSLTLOGLGISWLGSLNSRLSGALLIHNHTLVCVHT 479
Db 415 TDLAFELLETIRGRTRQHGQFSLAVVGLNTLSGLSLKSTISGDDVYISGNRLCIANT 474
QY 480 VPMDLFRNPQALHTANRPDECVGEGLAGHOLCARGHMGCGPTQVNCVSQFLGQOE 539
Db 475 INMKKLGFTPNQKTKINNNRAREKCKATNHYCNFLCSESGMGEPDPCVSCQVNSGRE 534
QY 540 CVEECRYLQGLPREYVNAHCLPCHPECOQONGSVTCGFPADQCVACAHKDPFCVAR 599
Db 535 CVDKCNILEGPRREFVESECIQHPCLPQTMNITCTGRRGPDNCIRKAHYVDGPHCVKT 594
QY 600 CPSCGVKDLSPYMPKPRDEGACQPCPINCSTHSCVLDKGC 642
Db 595 CPSCIMGBENNTL-VWKFADANNVCHLANCTYGCAGPGLKGC 636

RESULT 5
GOHUE
epidermal growth factor receptor precursor - human
N:Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB
C:Species: Homo sapiens (man)
C>Date: 15-Nov-1984 #sequence, revision 27-Nov-1985 #text, change 11-Jun-1999
C:Accession: A00641; A25772; S30024; A38672; A43615; A23062; A05281; A60143;
R:Ullrich, A.; Coussens, L.; Hayflick, J.S.; Dull, T.J.; Gray, A.; Tam, A.W.; Lee, J.
Tg, P.H.
Nature 309, 418-425, 1984
A:Title: Human epidermal growth factor receptor cDNA sequence and aberrant expression
A:Reference number: A00641; MUID:84219729; PMID:6328312
A:Accession: A00641
A:Molecule type: mRNA
A:Residues: 1-1210 <L>
A:Cross-references: EMBL:X00586; NID:q31113; PIDN:CAA25240.1; PID:q757924
A>Note: the authors translated the codon AAG for residue 540 as Asn
R:Ishii, S.; Xu, Y.; Stratton, R.H.; Roe, B.A.; Merlino, G.T.; Pastan, I.
Proc. Natl. Acad. Sci. U.S.A. 82, 4920-4924, 1985
A:Title: Characterization and sequence of the promoter region of the human epidermal
A:Reference number: A25772; MUID:85270436; PMID:2991899
A:Accession: A25772
A:Molecule type: DNA
A>Status: translation not shown
A:Residues: 1-29 <IS>
A:Cross-references: GB:M11234; NID:q181981; PIDN:AA52370.1; PID:q553272
R:Hailey, J.; Whitte, N.; Bennett, P.; Kitchingdon, D.; Ullrich, A.; Waterfield, M.
Oncogene Res. 1, 375-396, 1987
A:Title: The human EGF receptor gene: structure of the 110 kb locus and identificatio
A:Reference number: S30024; MUID:88217333; PMID:3329716
A:Accession: S30024
A:Molecule type: DNA
A:Residues: 1-29 <HA2>
A:Cross-references: EMBL:X06370; NID:q31118; PIDN:CAA29668.1; PID:q31119
R:Hailey, J.D.; Waterfield, M.D.
J. Biol. Chem. 266, 1746-1753, 1991
A:Title: Contributory effects of de Novo transcription and premature transcript termi
A:Reference number: A38672; MUID:91107677; PMID:1968448
A:Accession: A38672
A:Molecule type: DNA
A:Residues: 1-29 <HAL>
A:Cross-references: GB:M38425; NID:q181977; PIDN:AA63171.1; PID:q553271
A:Experimental source: carcinoma cell line A431-7
R:Xu, Y.; Ishii, S.; Clark, A.J.L.; Sullivan, M.; Wilson, R.K.; Ma, D.P.; Roe, B.A.;
Nature 309, 806-810, 1984
A:Title: Human epidermal growth factor receptor cDNA is homologous to a variety of RN
A:Reference number: A00642; MUID:84245835; PMID:6330563
A:Accession: A00642
A:Molecule type: mRNA
A:Residues: 'RCAMRRA', 150-187, 'KSYTQAV', 195, 'W', 197, 'A', 199-222, 'S', 224-304, 'RA', 307-
, '798-799', 'TD', '802-811', 'R', 813-942 <XUY>
A:Experimental source: A431 human carcinoma cells, which have large numbers of EGF re
R:Lin, C.R.; Chen, W.S.; Kuuliger, W.; Stolarsky, L.S.; Weber, W.; Evans, R.M.; Verna,

```

Science 224, 843-848, 1984  
A:Title: Expression cloning of human EGF receptor complementary DNA: gene amplification  
A:Reference number: A43615; MUID:84196372; PMID:6326261  
A:Accession: A43615  
A:Molecule type: mRNA  
A:Residues: 713-964 <LIN>  
A:Experimental source: epidermoid carcinoma cell line A431  
R:Slamen, F.A.; Gope, M.L.; Schulz, T.Z.; Wright, D.A.; Carpenter, G.; O'Malley, B.W.  
Biochem. Biophys. Res. Commun. 124, 125-132, 1984  
A:Reference number: A23062; MUID:85046483; PMID:6093780  
A:Accession: A23062  
A:Molecule type: mRNA  
A:Residues: 1028-1210 <STM>  
R:Weber, W.; Gill, G.N.; Speiss, J.  
Science 224, 294-297, 1984  
A:Reference number: A05281; MUID:84172183; PMID:6324343  
A:Accession: A05281  
A:Molecule type: protein  
A:Residues: 25-30, 'S', '32-51, 454-467 <WEB>  
R:Russo, M.W.; Lukas, T.J.; Cohen, S.; Staros, J.V.  
J. Biol. Chem. 260, 5205-5208, 1985  
A:Title: Identification of residues in the nucleotide binding site of the epidermal growth factor receptor  
A:Reference number: A60143; MUID:85182650; PMID:2985580  
A:Accession: A60143  
A:Molecule type: protein  
A:Residues: 740-744, 'X', '746-747 <RUS>  
R:Motczkowski, B.; Mosig, G.; Cohen, S.  
Nature 309, 270-273, 1984  
A:Title: ATP-stimulated interaction between epidermal growth factor receptor and superoxide dismutase  
A:Reference number: A38023; MUID:84191554; PMID:6325948  
A:Accession: A38023  
A:Contents: annotation; receptor activity  
A:Note: the EGF receptor (and other tyrosine kinases) can nick double-stranded DNA  
R:Chen, W.S.; Lazar, C.S.; Lund, K.A.; Welsh, J.B.; Chang, C.P.; Walton, G.M.; Der, C.J.  
Cell 59, 33-43, 1989  
A:Title: Functional independence of the epidermal growth factor receptor from a domain I  
A:Reference number: A33331; MUID:90003233; PMID:2790960  
A:Accession: A33331  
A:Contents: annotation; internalization signal  
C:Comment: Binding of EGF to the receptor leads to internalization of the EGF-receptor complex  
C:Genetics:  
A:Gene: GDB:EGFR  
A:Cross-references: GDB:120610; OMIM:131550  
A:Map position: 7p12.3-7p12.1  
A:Superfamily: epidermal growth factor receptor; protein kinase homology  
C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phosphatase; signal transduction; status predicted <SIG>  
F:1-24/Domain: signal sequence #status predicted <SIG>  
F:25-1210/Product: EGF receptor #status predicted <MAT>  
F:25-645/Domain: extracellular #status predicted <EXT>  
F:75-300/Domain: EGF receptor extracellular domain repeat <EE1>  
F:390-600/Domain: EGF receptor extracellular domain repeat <EE2>  
F:668-1210/Domain: transmembrane #status predicted <TM>  
F:668-1210/Domain: intracellular #status predicted <INT>  
F:710-915/Domain: protein kinase homology <KIN>  
F:718-726/Region: protein kinase ATP-binding motif  
F:999-1046/Region: coated-pit mediated internalization signal  
F:1047-1210/Region: inhibitory  
F:128, 175, 352, 413, 444, 528, 603/binding site: carbohydrate (Asn) (covalent) #status predicted  
F:745/Active site: lys #status experimental

Query Match 38.8%; Score 1534; DB 1; Length 1210;  
Best Local Similarity 42.9%; Pred. No. 1,66-94;  
Matches 283; Conservative 109; Mismatches 240; Indels 28; Gaps 6;

Qy 11 LLLALLPAGAA--SYOVCTGDMKRLPASPETHIDMRHLHYOCGOVQVQLNLELTPN 68  
Db 14 LLAALCPSPRALLEKKVQCGTSNKLTQIGTFEDHFLSLQRFMNCDEVVLNLELTYVGRN 73  
Qy 69 ASLSFIDIOEVQGVYVLAHNOVROVPLQRLRIYVGTQLFEDNYALAVLNGDPLNNTTP 128  
Db 74 YDLSTFKIQVAGVIALMTVERIPLENQIIRGNMYENSAVALVSTND----- 126  
Qy 129 VTGASPGGLRELQRLSTLEILKGVLIQNRDOLCYDTIIMKIDFHKNNQALATLITDNR 188  
Db 127 ---ANKTGKLELPMKNIQELIHGAVRFESNPNALCNVESIQNRDIYSDPLSNMSMDQNH 183

Qy 189 SRACHPCSPMGKSGRCWKESSSDQSLRTVCAGGCA-RCKGPIPTDCHEQCAAGCTGP 247  
Db 184 LGSCKQCDPSCPNCGSGWAGEBNCOKLTKITICAQCCSGCRKRSDDCHQCAAGCTGP 243  
Qy 248 KHSDDLACLNHNSGICELHCPALVTYNTDFESPNEGRRTGASCVTACPYNLSTD 307  
Db 244 RESDCLVCRKRDEATCDTQCPPLMLTYPTTYQMDVNEBGRKSPATCYKCKPRYVYTD 303  
Qy 308 VGSCTLVCPNLNQEYTAEDGTORCEKCSKPCARVCYGLGMEHLREVAVTSANIOEFAG 367  
Db 304 HGSCVACGADSYEM-EEGVYRCKCKCEPCRVKVCNGIGIEFKDLSINATNIKHFKNK 362  
Qy 368 KIRGSFLAPDESFPDPSANTAPLQPPOLQVETLEITGYLISAMPDPLDLSVQN 427  
Db 363 TSIISGDLHLPLVAFRGDSFTHTPPIDPDELILKVKKEITGFLIQAMPENRTDLHAEN 422  
Qy 428 LOVIRGRLIHNGAVSLTLOGISWLGRLSRELGGSLALHNTLHCFVTPMDOLF 487  
Db 423 LEITGRKQHGQPSLAVNSLITSLGRLSKLEISDGVITISGNKNTCYANTIMKKLFG 482  
Qy 488 NPHQALLTANRPDECEVGEGLACHQLCARHCWGPPTQVCNSQFLRGQCEVECHYL 547  
Db 483 TSGQRTKIISNRGENSKATGQVCHALCSPEGCMGPEPRDCVSCRNVRGRRECVDKL 542  
Qy 548 OGPREYVNAHRCPLCHPECPQNSVTCFPEADQCAAHYKDPFCVACRPGCVAPD 607  
Db 543 EGEPREFENSECICQHCCELPQAMNITCTGRGPNCTOCAGHYIDGRCVCTCPAGVGE 602  
Qy 608 LSYMPIMKPEDEGACQCPINCFHSCVLDLDCGCPAPGRASPLTSQNEDEGPASPPLST 667  
Db 603 NNTL-VMKYADAGVCHLCHNCNTYGCYGLGEGPTN-----GPKIPSIAT 648

RESULT 6  
Epidermal growth factor receptor precursor - mouse  
A:Species: Mus musculus (house mouse)  
C:Date: 06-Jan-1995 #sequence revision 06-Jan-1995 #text change 18-Jun-1999  
C:Accession: A53183; A43818; S24942; A28941; S45325; I49643  
R:Luetkeke, N.C.; Phillips, H.K.; Qiu, T.H.; Copeland, N.G.; Earp, H.S.; Jenkins, N.A.  
Genes Dev. 8, 399-413, 1994  
A:Title: The mouse waved-2 phenotype results from a point mutation in the EGF receptor  
A:Reference number: A53183; MUID:94170986; PMID:8125255  
A:Accession: A53183  
A:Molecule type: mRNA  
A:Residues: 1-1210 <LUE>  
A:Cross-references: GB:U03425  
R:Aviv, A.; Iax, I.; Ulrich, A.; Schlessinger, J.; Givol, D.; Morse, B.  
Oncogene 6, 673-676, 1991  
A:Title: Comparison of EGF receptor sequences as a guide to study the ligand binding  
A:Reference number: A43818; MUID:91232866; PMID:2030916  
A:Accession: A43818  
A:Molecule type: mRNA  
A:Residues: 1-714 <AVI>  
A:Cross-references: GB:X59698  
R:Eisinger, D.P.; Serrero, G.  
Submitted to the EMBL data library, June 1992  
A:Reference number: S24942  
A:Accession: S24942  
A:Molecule type: mRNA  
A:Residues: 969-971, 'K', '973-1115, 'D' <EIS>  
A:Cross-references: EMBL:Z12608  
R:Heisermann, G.U.; Gill, G.N.  
J. Biol. Chem. 265, 13152-13158, 1990  
A:Title: Epidermal growth factor receptor threonine and serine residues phosphorylate  
A:Reference number: A28941; MUID:88330814; PMID:3138233  
A:Accession: A28941  
A:Molecule type: protein  
A:Residues: 689-694, 'X', '696-704, 'V', '706-707, '989-992, 'XX', '995-996, 'X', '998-1000, '1002-10  
R:Riggs, M.L.; Dunn, A.R.; Alexander, W.S.  
Submitted to the EMBL data library, April 1994  
A:Description: The complete cDNA sequence of the Mouse Epidermal Growth Factor Recept

A:Reference number: S45325  
 A:Accession: S45325  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1971, 'K', 973-1210 <VER>  
 A:Cross-references: EMBL:X78987; NID:g488830; PIDN:CAA5587.1; PID:g488831  
 R:Parla, B.C.; Das, S.K.; Andrews, G.K.; Day, S.K.  
 Proc. Natl. Acad. Sci. U.S.A. 90, 55-59, 1993  
 A:Title: Expression of the epidermal growth factor receptor gene is regulated in mouse  
 A:Reference number: I49643; MIMD:93126380; PMID:7678348  
 A:Accession: I49643  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 12-20, 22-132 <RES>  
 A:Cross-references: GB:I06864; NID:g193001; PIDN:AA53029.1; PID:g567201  
 C:Genetics:  
 A:Gene: EGFR  
 C:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C:Keywords: ATP; growth factor receptor; kinase-related transforming protein; phosphop  
 F:1-24/Domain: signal sequence #status predicted <SIG>  
 F:648-670/Domain: transmembrane #status predicted <TM>  
 F:712-977/Domain: protein kinase homology <KIN>  
 F:720-728/Region: protein kinase ATP-binding motif  
 F:680,695/Binding site: phosphate (Thr) (covalent) #status experimental  
 F:697,1070,1071/Binding site: phosphate (Ser) (covalent) #status experimental  
 F:993/Binding site: (or 997) phosphate (Ser) (covalent) #status experimental  
 F:1028/Binding site: (or 1030 or 1032) phosphate (Ser) (covalent) #status experimental  
 F:1197/Binding site: phosphate (Tyr) (covalent) #status experimental

Query Match 38.8%; Score 1533.5; DB 2; Length 1210;  
 Best Local Similarity 44.4%; Pred. No. 1.8e-94;  
 Matches 284; Conservative 99; Mismatches 233; Indels 23; Gaps 7;

QY 11 LLLALLPPGA--STOYCTGDMKRLPASPETHLMDLRHLHYGCCOYVGGLEITYPTN 68  
 Db 14 LTRALCAAGALEERKVCQCGSNRLTQLGFEDHFLSLQRYNNCEVVLGLETYYQNR 73  
 QY 69 ASLSFLDIOEVQGYVLAHQVROVPLQRLRIYRGTFQEDNALVALVDGDEPLNNTTP 128  
 Db 74 YDLSFLTIQEVAGYVLAINTVERIPLENQIIRGNALVENTALALISN----- 124  
 QY 129 VTGASPEGRLQRLSLTEILKGGVLIQRNPQLCYDTILMKDI----FKNNQALATLI 184  
 Db 125 -YGNRTGLRELPMRNQETILIGAVRSPNPILCNMDTIQMRDIQVNVFMSNMSMDT--- 180  
 QY 185 DTNRSRACHCSPKCSKRCMGESSEDCQSLTRVYAGGCA-RKQGLPLPDCCHQCAAG 243  
 Db 181 -QSPSSCPKCDPSCPNMGSCWGGEGCOKLTKICAOQCSHRCGRSPSCCHNCCAAG 239  
 QY 244 CTGFRHSDCLACLFHNSGICELCPALVTYNTDFESMPREGRYFGASCYACPEYNY 303  
 Db 240 CTGFRSDCLVCKQFOEATCKDTCPLMLYNPITYQMDVNPBEKTYFGATCVAKCRNY 299  
 QY 304 LSTDVGSCTLVCPLAHQEVTAEADGTQRCCKSPKCANVYGLGMEHLREYRAVTSANIOE 363  
 Db 300 VVVDHGCVACVAGDYEV--EEDGIRCKKCDKPCFRKVCNGIGEFKEDTLISATNAIKH 358  
 QY 364 FAGCKRTFGSLAPLPESFDDPASNTAPLQPEQLVFEETLEETGYLISAMPDLSLDLS 423  
 Db 359 FKCTALISGLHLIPVAFKDSFTRPPLDRELLEIKTYKETITGILLIADMPDNMDLH 418  
 QY 424 VFONLQVIRIRLHNGAVSTLQGLISWLGSLRLRELSGLALIHNNTHLCFYHVPMD 483  
 Db 419 AFENLEIRRTQHGOFSLAVYGLNTSLGRSLKEISBDVITISGRNRCYANTINMK 478  
 QY 484 QLFERNPHOALLHTANPEDECVGBGLACHQLCARGHCWGPGPTQCVNCSQFLRGQECVEE 543  
 Db 479 KLRTGPMQKTKIMNNRAEKDKAVNHVCNPLCSSEGCMGEPBRCVSCQNVNRSRRECEV 538  
 QY 544 CRLVGLGPREYVANAHLPLCHPEQCPONGSVYTCGFPADOCVACAHKKDPFCYARPSG 603  
 Db 539 CNLGEPRFEVENSECIGCHPECLPQAMNITCTGRRGPDNDCIOCAHYIDGHCVKTCGPA 598

QY 604 VKPDLSTYPIWKPEDEBACQPCPINCTHSCVLDLDDKGC 642  
 Db 599 IMGENNRL-VMKYADANNVCHLCHANCTYCAGAGLQGC 636

# RESULT 7

## tychilv

epidermal growth factor receptor precursor - chicken

N:Contin: protein-tyrosine kinase (EC 2.7.1.112) erbB

C:Species: Gallus gallus (chicken)

C:Date: 28-Feb-1986 #sequence\_revision 05-May-1995 #extL\_change 04-Feb-2000

C:Accession: A27720; A00643

R:Lax, I.; Johnson, A.; Hawk, R.; Sap, J.; Bellot, F.; Winkler, M.; Ullrich, A.; Venn

Mol. Cell. Biol. 8, 1970-1978, 1988

A:Title: Chicken epidermal growth factor (EGF) receptor: cDNA cloning, expression in

A:Reference number: A27720; MIMD:88261272; PMID:3260329

A:Accession: A27720

A:Molecule type: mRNA

A:Residues: 1-1223 <LAX>

A:Cross-references: GB:M20386

R:Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines

Cell 41, 719-726, 1985

A:Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and

A:Reference number: A00643; MIMD:85228222; PMID:298784

A:Accession: A00643

A:Molecule type: mRNA

A:Residues: 585-1223 <NTL>

A:Cross-references: GB:M10066

C:Genetics:

A:Gene: erbB

C:Superfamily: epidermal growth factor receptor; protein kinase homology

C:Keywords: alternative splicing; ATP; autophosphorylation; glycoprotein; growth fact

peptic protein kinase

F:1-30/Domain: signal sequence #status predicted <SIG>

F:31-123/Domain: epidermal growth factor receptor #status predicted <MAT>

F:31-654/Domain: extracellular #status predicted <EX>

F:81-307/Domain: EGF receptor extracellular domain repeat <BE1>

F:397-610/Domain: EGF receptor extracellular domain repeat <EE2>

F:655-677/Domain: transmembrane #status predicted <TM>

F:678-1223/Domain: intracellular #status predicted <INT>

F:719-984/Domain: protein kinase homology <KIN>

F:727-735/Region: protein kinase ATP-binding motif

F:136,202,280,361,370,422,575,580,615,635/Binding site: carbohydrate (Thr) (covalent)

F:192,650/Binding site: carbohydrate (Ser) (covalent) #status predicted

F:687/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicte

F:754/Active site: Lys #status predicted

F:1100,1183,1208/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #

Query Match 38.2%; Score 1509.5; DB 1; Length 1223;  
 Best Local Similarity 44.2%; Pred. No. 7.2e-93;  
 Matches 286; Conservative 103; Mismatches 235; Indels 23; Gaps 9;

QY 8 RWGILLALLPPGA-----STOYCTGDMKRLPASPETHLMDLRHLHYGCCOYVGGNLE 61  
 Db 13 RGAIVELVLLGLVAGLCAVEERKVCQCGSNRLTQLGFEDHFTSLQRYNNCEVVLGLETYYQNR 72  
 QY 62 LTYLPNANSLSPLODIOEVQGYVLAHQVROVPLQRLRIYRGTFQEDNALVALVDGDEPLNNTTP 128  
 Db 73 ITVEHNRDLFTKTIQEVAGYVLAINTVERIPLENQIIRGNALVENTALALISN----- 124  
 QY 122 PLNNTPTVTGASPGRLQRLSLTEILKGGVLIQRNPQLCYDTILMKDI----FKNNQALATLI 181  
 Db 133 -MKNTQ-----GLRELPMKRLSELNCGVKSINNPRIKCNMDVTYVLANNDIITDSRK-PL 182  
 QY 182 TLID-TNRSRACHCSPKCSKRCMGESSEDCQSLTRVYAGGCA-RKQGLPLPDCCHQ 239  
 Db 183 TVLDFASNLSPKCHPNCNTEDEHMGAGEONCOTLTFKVICAOQCSHRCGRSPSCCHNCCAAG 242  
 QY 240 CAAGCTGPRKSDCLACLFHNSGICELCPALVTYNTDFESMPREGRYFGASCYACPEYNY 303  
 Db 243 CAAGCTGPRSDCLACLFHNSGICELCPALVTYNTDFESMPREGRYFGASCYACPEYNY 302  
 QY 300 PYNLTSTVGSCTLVCPLAHQEVTAEADGTQRCCKSPKCANVYGLGMEHLREYRAVTSANIOE 363



```

Db      303 PHNYVYDTHGSCVSRCSNDITYEV -BENGVRCKCKKDDGLCSKVCNIGIGELKGIISTANT 361
QY      360 NIOEFAGCKKJFGSLAFIPESFDGDPASNTAPLOEOLQVFEETLEITGYLISAMPDSL 419
Db      362 NIDSEKNTKTINGDVSIPLVAFIADAFKTLPLDPKRLDVPRTYKEISGFLIIQAMPDWA 421
QY      420 PDLSTVFOMLQVIRGILHLNGAVSLTLOGLSIWSGLBSRLRELGSSLALIHNTILCFVHT 479
Db      422 TDLVAFENLEIRGTRKHOGVSLAVVNLKIOSLGLRLKELISDDIDIAIMKKNLCYADT 481
QY      480 VPMOLFENPQALLHTNANREDECEVGEGLACHOLCARHGHCMPGPTQVNCNSOPLRGOE 539
Db      482 MNMWSLFSATQSOQKRIIONRKKNKDTADRHVCDPLCSVDYGCMBPFPFCFSRFRFSROKE 541
QY      540 CVEBCRVLOGLPREYVNAHRHCLPCHPECOPONG---SVTCFGRPEADQCVACAHYKDPFC 596
Db      542 CVKONCIIIOGEPREFERSKCLPCHSECTLYONSTAYNTTCSGPGDHCMAHFIDGBHC 601
QY      597 VARPSGKPRDLSYMPIMKFPDEBGAQCPRLNCHSCVDDDDKCCP 643
Db      602 VKACPAGVLENDTL-VKVTADANAAVQLCHPNCRTGCKGPGLEBCP 647

```

RESULT 8  
A47253  
Epidermal growth factor receptor, HER4 - human  
C:Species: Homo sapiens (man)  
C:Date: 22-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 18-Jun-1999  
C:Accession: A47253  
R:Plowman, G.D.; Culouscou, J.M.; Whitney, G.S.; Green, J.M.; Carlton, G.W.; Foy, L.; N  
Proc. Natl. Acad. Sci. U.S.A. 90, 1746-1750, 1993  
A:Title: Ligand-specific activation of HER4/p180erbB4, a fourth member of the epidermal  
A:Reference number: A47253; MUID:93189574; PMID:8383326  
A:Accession: A47253  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: nucleic acid  
A:Residues: 1-1308 <P/O>  
A:Cross-references: GB:I07868; NID:g337359; PIDN:AAB5946.1; PID:g337360  
A:Note: sequence extracted from NCBI backbone (NCBI:P:126842)  
C:Superfamily: epidermal growth factor receptor; protein kinase homology  
C:Keywords: ATP; growth factor receptor  
F:716-981/Domain: protein kinase homology <KIN>  
F:724-732/Region: protein kinase ATP-binding motif

[illegible][illegible]

RESULT 9  
A36223  
kinase-related transforming protein (erbB3) (EC 2.7.1.-) precursor - human  
C:Species: Homo sapiens (man)  
C:Date: 04-Oct-1991 #sequence\_revision 13-Jan-1993 #text\_change 17-Nov-2000  
C:Accession: A36223; 159164  
R:Krus, M.H.; Issing, W.; Miki, T.; Popescu, N.C.; Aaronson, S.A.  
Proc. Natl. Acad. Sci. U.S.A. 86, 9193-9197, 1989  
A:Title: Isolation and characterization of ERBB3, a third member of the ERBB/epidermal  
A:Reference number: A36223; MUID:90083234; PMID:2687875  
A:Accession: A36223  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1342 <RKA>  
A:Cross-references: GB:M29366  
R:Plowman, G.D.; Whitney, G.S.; Neubauer, M.G.; Green, J.M.; McDonald, V.L.; Todaro,  
Proc. Natl. Acad. Sci. U.S.A. 87, 4905-4909, 1990  
A:Title: Molecular cloning and expression of another epidermal growth factor receptor  
A:Reference number: 159164; MUID:90311312; PMID:2164210  
A:Accession: 159164  
A:Status: preliminary; translated from GB/EMBL/DBD  
A:Molecule type: mRNA  
A:Residues: 1-559,'G', 561-957,'F', 959-1063,'G', 1065-1342 <RES>  
A:Cross-references: GB:M34309; NID:g183990; PIDN:AAA35979.1; PID:g306841  
C:Genetics:  
A:Gene: GDB:ERBB3; HER3  
A:Cross-references: GDB:119880; OMIM:190151  
A:Map position: 12q13-12q13  
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase hom  
C:Keywords: ATP, phosphotransferase  
F:707-972/Domain: protein kinase homology <KTN>  
F:715-723/Region: protein kinase ATP-binding motif

	Query Match	36.3%	Score 1435;	DB 2:	Length 1342;	
	Best Local Similarity	43.6%;	Pred. No.7.5e-88;			
	Matches 278; Conservative	92;	Mismatches 244;	Indels 24;	Gaps 12;	
QY	10 GLLALLPPGAA--STVCTGTDKMLRLPASPEHLEIDMLRHLYOGCCVVGNLEETLPT	67	:     :	:	:	
Db	11 GLLFSLARGSEVGNQAOCPTGLNLSTVGAEHQYTLYKLKYERCEVVMNLEIVLGH	70				
QY	68 NASLSFLDDIQEVSQVYLIAHNQVRVPLOLRIRYRGTQLTEEDNTALAVLDNGPLNNTT	127				
Db	71 NADISFLQIRIVGTGYLVAMNEFSTLPPLNLRVRRGRQVDGKRAIVM-----LNNT	125				
QY	128 PVTGASPGELREIOLRSLETELKGVLIQRNPDLCOYTIILMKDIEFHKNNOIALTLIDTN	187				
Db	126 ----NSSHALNQRLQTGLEITLISGSVYLEKNDKLCMHMTIDMRDIWRD---AIIYKD	178				
QY	188 RSRACHPSPMCKSGRCMGESSEDQCILTRIVCAGGC--ARKGPLPDCCHBOCAAGTG	246				
Db	179 NGRSPCPCHCYKC--RCMGPGSPEDQTYLTIKCIAQCCHGCFGPMPNCCCHDECAGCGSG	237				

[illegible]

RESULT 10  
epidermal growth factor receptor homolog precursor - rat  
JC4387  
N:Alternate names: ErbB3 protein; HER3 protein  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 17-Jan-1996 #sequence\_revision 19-Apr-1996 #text\_change 13-Nov-1998  
C:Accession: JC4387  
R:Hellier, N.J.; Kim, H.H.; Greaves, C.H.; Sierke, S.L.; Koland, J.G.  
Gene 165, 279-284, 1995  
A:Title: Cloning of the rat ErbB3 cDNA and characterization of the recombinant protein.  
A:Reference number: JC4387; MUID:96096535; PMID:8522190  
A:Accession: JC4387  
A:Molecule type: mRNA  
A:Residues: 1-1339 <HEL>  
A:Cross-references: GB:U29339; NID:g915389; PID:g915390  
A:Experimental source: liver  
A:Note: The authors translated the codon AAC for residue 369 as Thr and GTr for residue  
C:Comment: This protein is a functional heregulin receptor that transduces signals to th  
C:Genetics:  
A:Gene: ErbB3  
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolo  
C:Keywords: ATP; growth factor receptor; liver; phosphoprotein; transmembrane protein  
F:1-19/Domain: signal sequence #status predicted <SIG>  
F:20-1339/Product: epidermal growth factor homolog #status predicted <NAT>  
F:640-659/Domain: transmembrane #status predicted <TM>  
F:705-970/Domain: protein kinase homology <KIN>  
F:713-721/Region: protein kinase ATP-binding motif  
F:939,1051,1156,1194,1196,1219,1257,1259,1273,1286,1325/Binding site: phosphate (Tyr) (G

Query Match	34.2%	Score 1351.5	DB 2	Length 1339
Best Local Similarity	42.2%	Pred. No. 2.8e-82		
Matches 269; Conservative	87;	Mismatches 253;	Indels 29;	Gaps 13;

[illegible][illegible]

RESULT 11  
 S06142  
 protein-tyrosine kinase (EC 2.7.1.112) mrk-y precursor - southern platyfish  
 N:Alternate names: epidermal growth factor receptor homolog; kinase-related  
 C:Species: Xiphophorus maculatus (southern platyfish)  
 C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 18-Feb-2000  
 C:Accession: S06142; S13809  
 R:Witlbrodt, J.; Adam, D.; Maltischek, B.; Maeucler, W.; Raulf, F.; Telling, A.; Robe  
 Nature 341, 415-421, 1989  
 A:Title: Novel putative receptor tyrosine kinase encoded by the melanoma-inducing Tu  
 A:Reference number: S06142; MUID:90015140; PMID:2797166  
 A:Accession: S06142  
 A:Molecule type: DNA  
 A:Residues: 1-1166 <MT>  
 A:Cross-references: EMBL:X16891; NID:965290; PIDN:CMA34770.1; PID:965291  
 R:Adam, D.; Maeucler, W.; Scharlt, M.  
 Oncogene 6, 73-80, 1991  
 A:Title: Transcriptional activation of the melanoma inducing xmrk oncogene in xiphoph  
 A:Reference number: S13807; MUID:91125882; PMID:1846957  
 A:Accession: S13809  
 A:Status: preliminary; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 821-1025, 'N', 1027-1098, 'A', 1100-1166 <ADA>  
 A:Cross-references: EMBL:X56319; NID:965284; PIDN:CMA39763.1; PID:965285  
 C:Genetics:

A:Gene: mirk  
A:Map position: Y  
A:Introns: 872/3: 898/1: 947/1: 979/3: 1025/3: 1056/1  
C:Superfamily: epidermal growth factor receptor; protein kinase homology  
C:Keywords: ATP; growth factor receptor; phosphotransferase; transmembrane protein; t  
F:1-25/Domain: signal sequence #status predicted <SIG>  
F:26-1166/Product: kinase-related transforming protein (Tu) #status predicted <MAT>  
F:707-972/Domain: protein kinase homology <KIN>  
F:715-723/Region: protein kinase ATP-binding motif

Query Match 33.2%; Score 1313.5; DB 1; Length 1166;  
Best Local Similarity 41.6%; Pred. NO. 8.3e-08;  
Matches 268; Conservative 93; Mismatches 259; Indels 25; Gaps 12;



```

0Y 256 LHNHNSGICELXCPALVYNTDTEESMNPBEGRYEFGASCTACVYNTLSDVSGCTLYC 315
Db 274 KNFDEAVSKCECPRMKRYNPTTYLTETNPBCKAYGATVKECP-GHLNLDGAVRSC 332
0Y 316 PLHNOEVTAAEDTQCEKCSKPCARVCYGLGMEHLREVRAYTSANIOEPAGCKKJFGSLA 375
Db 333 FQDKKDKGGE-----CVPONGCPKTCPGVYVLH-----AGNIDSFRCNTVIDGNIR 379
0Y 376 FLPESEFDG--DPAWNTA-----PLQPELOLVETLEETITGYLISAMPDLSLDLVSQN 427
Db 380 ILDOFFSGQDYYANTWGPRIYPLDPERREVEFSTVKITGYLNEGTHPOFRNLSTYERN 439
0Y 428 LQVYGRILHNGAY-SLTLOGIGISMLGRSLRETSGLALHNHTHLCFYHTVWDOLF 486
Db 440 LETHGRQMESMFALAIYKSSLSLEMRNLKOTSSSVYIQHNRDLCYYSNIRMPAIQ 499
0Y 487 RNPQOALLHTANRPDECEVGBGLACHOLCARGHCWGPPTQCVNCSQFLRQOEVEECRY 546
Db 500 KEPEKQVWVWENIRLADLEKNKNTICSDOCNEDGCAGCATDCLTCKNPNFNNGTCLADGY 559
0Y 547 LOGLPREYVNAHCLRCPEQPONGSWTQCFEPREDDQVCAAHYKDPPEFCVARGCSGKRP 606
Db 560 ISNAVK--FDNFTTCIICPECR-----TCNGAGDHQOECHVHRDGOCHGVSECKN-- 608
0Y 607 DLSYMPIMKFPDEBACOPCPINCTHSCVDLDKCAPAEORASPLTSONEDLGVASPL- 664
Db 609 -KYND-RGVCRCCHATC-----DGC-----TGPKDITIGIACATTC 641
0Y 665 -----DSTFYRSLLEDDMGMDLVDABEELYVPO-QGFRCF 697
Db 642 NLAITNNDAIVKRCCLKDKCPDG-Y-FWEYVHPQOGSSLKRP 681

```

RESULT 14  
E888257  
protein let-23 [imported] - Caenorhabditis elegans  
C.Species: Caenorhabditis elegans  
C.Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 24-Aug-2001  
C.Accession: E888257  
R.anonymous, The C. elegans Sequencing Consortium.  
Science 282, 2012-2018, 1998  
A.Title: Genome sequence of the nematode C. elegans: a platform for investigating biology  
A.Reference number: A75000; MUID:99056613; PMID:9851916  
A.Note: See websites genome.wustl.edu/gsc/C\_elegans/ and www.sanger.ac.uk/Projects/C\_elegans/  
A.Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and  
A.Accession: E888257  
A.Status: Preliminary  
A.Molecule type: DNA  
A.Residues: 1-1323 <STO>  
A.Cross-references: GB:chr\_II; PIDN:CAA93882.1; PID:g3881523; GSPDB:GN00020  
C.Genetics:  
A.Gene: let-23  
A.Map position: 2  
C.Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homold

```

Query Match      18.2% Score 718.5 DB 2 Length 123;
Best Local Similarity 28.0%; Pred. No. 4,9e+40;
Matches 191, Conservative 107, Mismatches 256, Indels 129, Gaps 28;

OY   25 VCTGDMKRLPASPETHLDMLRHLYGGCGVQGNLELTYPTN-----        68
       :::|||          ::|||:: |||::|||:: |
Db    39 ICSGTNGISRYXTGNI-LEDLETMYRGCRARYGMLETITWEANIKKRRESTNSTVDPK 97
               -
OY     69 -----ASTSLFLODIOEVGYVLIAHNOVRQPFLORLRIVRSTGFEDNYALAVADNDGP 122
           ::|||::|||::|::|::|::|::|::|::|::|::|::|::|::|::|
Db     98 NEDSPLKSINFDNLDELIRGLIIYRANIQTSPFRLLRVYGDEVFHDN-ALIHKNNK- 155
                               :::|::|::|::|::|::|::|::|::|::|::|
OY    123 LNTTPYVGASGGRLEQLRSLTTELINGGVLIQNPOLCY-QDTILMKDLIFHKNNQAL 181
                ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db    156 -----VHEVMRELVIIRNGSVTIQNPCKCYIGDKIDMKLELYDPD--VQ 199
                                :::|::|::|::|::|::|::|::|::|::|::|
OY   182 FLIDNRSRACH-----PCSPMGKSCRWGESSEDCOSLTRFYCACAGCARC---KGPL 231
         : || ::||| : ||| ||| ||| : |||
```

```

Db 260 KVEIINSHQHCYNKSKMAKHCESC -NDKMGSSDDNQCQRYYRBSVPCRSQSCPCYNSSTS 258
OY 222 PFDCCHEOCAGCTGPKRKSDDCLACLFHFNHSGICELHCPALVYTYTDTFFESPMPNEGKTYF 291
Db 259 SYECCDSACLOGCGHGRKHCACSKYELDGCICLETCPSRKIFHNHKTGRVLVFNPDGRYQN 318
OY 292 GASCYTAAPRYNTL-STVDSGCTIWC-PLHNOEVALEDOGTQCEKC-SKPCARVCYGGAME 348
Db 319 GNHCYKCEPPELLIENDV--CVRHCSODGHNTDAIKD--VRECEKCRSSSCKICTYVG-- 372
OY 349 HIREVRAVTSANIOEFACCKKIFGSLAFLEPSEFDGDPASNTAPLQOPDOLVFEETLEITG 408
Db 373 HU-----TNETIKKLECEQIDGHL-IEHAF-----TYEQLKLEFVKLVSE 414
OY 409 XLYISAMPDLSPLDSVPNOLOVIRGRLIHGNAYSILT-OGLGISMLGIRSLRETSGLAL 467
Db 415 YITTY--QOONYDLKFLKLNLOIIEGRKLIHNRMALAIYIQCDDELISNLSKLIKITAVAL 472
OY 468 IHHNHTLCFVHTVPDOLF---RNPQALHTGTANRPEDECVCBGLACHOLCARGHCWGP 523
Db 473 IMKNHRLCYVSKIDWSSIITSKGDKNRPSLIAENBDSKICLETBQVPODKMCKNRKCGWK 532
OY 524 GPLOCVNSOFLROGECEYECRCVLOGLPREYVNAHRLCPHRECOQPNONGSYTCGPRADQ 583
Db 533 EPEDLECKTWSVGTCEYKCDT--KGLRNQTSMK-CERCSPBEE-----TCNGLGLELD 584
OY 584 CVACAH-----YKDRPFVCARQPSGVKPDLSYMPILKPFDEBAGACQCPINCTHSCVD 636
Db 585 CLTCHKTLTNSDFGNRMCEVHDC-----PVSHFTQKNVCEKCHPTCY----- 628
OY 637 LDDKCGRAEQRASPLTSONEDLG 659
Db 629 --DNGC-----TGPDSNLG 640

```

RESULT 15  
 S70712  
 protein-tyrosine kinase (EC 2.7.1.112) let-23 precursor - *Caenorhabditis elegans*  
 N:Alternate names: receptor tyrosine kinase let-23  
 C:Species: *Caenorhabditis elegans*  
 C:Date: 21-Apr-1997 #sequence\_revision 09-May-1997 #text\_change 11-Jan-2002  
 C:Accession: S70712; S73101; S13422; T27682  
 R:Sakal, T.; Koga, M.; Ohshima, Y.  
 J.Mol. Biol. 256, 548-555, 1996  
 A:Title: Genomic structure and 5' regulatory regions of the let-23 gene in the nemato  
 A:Reference number: S70712; MUID:96177760; PMID:8604137  
 A:Accession: S70712  
 A:Status: nucleic acid sequence not shown  
 A:Molecule type: DNA  
 A:Residues: 1-1374 <SAK>  
 A:Cross-references: EMBL:D63426  
 A:Experimental source: Strain N2  
 R:Koga, M.  
 submitted to the EMBL Data Library, July 1995  
 A:Reference number: S73101  
 A:Accession: S73101  
 A:Molecule type: DNA  
 A:Residues: 1-50, 'G', 52-1374 <KOG>  
 A:Cross-references: EMBL:D63426; NID:g1407562; PIDN:BAA09729.1; PID:g1407563  
 A:Experimental source: strain N2  
 R:Arlian, R.V.; Koga, M.; Mendel, J.E.; Ohshima, Y.; Sternberg, P.W.  
 Nature 348, 693-699, 1990  
 A:Title: The let-23 gene necessary for *Caenorhabditis elegans* vulval induction encode  
 A:Reference number: S13422; MUID:91080919; PMID:1979659  
 A:Accession: S13422  
 A:Molecule type: mRNA  
 A:Residues: 52-1374 <ARO>  
 R:Thomas, K.  
 submitted to the EMBL Data Library, March 1996  
 A:Reference number: Z20404  
 A:Accession: T27682  
 A:Status: preliminary; translated from GH/EMBL/DBJ  
 A:Molecule type: DNA

A:Residues: 52-1374 <WII>  
A:Cross-references: EMBL:Z70038, PTDN:CAA93882.1; GSPDB:GN00020; CESP:ZK1067.1  
A:Experimental source: clone ZK1067  
C:Genetics:  
A:Gene: let-23; CESP:ZK1067.1  
A:Map position: 2  
A:Map coordinates: 44/1; 90/1; 112/3; 165/3; 217/1; 290/1; 379/1; 418/1; 448/2; 565/1; 608/1  
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homoloid  
C:Keywords: ATP; phosphotransferase; transmembrane protein; tyrosine-specific protein kinase  
F:1-28/Domains: signal sequence #status predicted <SIG>  
F:29-1374/Product: protein-tyrosine kinase let-23 status predicted <MNT>  
F:94-1199/Domains: protein kinase homology <KIN>  
F:942-950/Region: protein kinase ATP-binding motif

Query Match 18.2%; Score 718.5; DB 2; Length 1374;  
Best Local Similarity 28.0%; Pred. No. 5,16-40;  
Matches 191; Conservative 107; Mismatches 256; Indels 129; Gaps 28;

QY 25 VCTGDMRLRPASETHLMDRIHLHYGCCVYQGLLETLPTN----- 68  
DB 90 LCSGTTNGISRYXTGNI-LEDELTMYRCGRRYGNGLETLTWEANEIKKMRSTNSTYDPR 148  
QY 69 -----ASLSEFLDIDIOEVGYVLIHANOVOYPLORLRIYVGTOLFEDNYRLAVLNDGP 122  
DB 149 NEDSLKAINFDNIEELRSGSIITIRANIQKISPRRLVYIGDEVFHDN-ALYIHKNDK- 206  
QY 123 LNNTPVYTGASPGGLRELDQLSLTEILKGVLIQHPOLCY-ODYTIHKDIFHKNNQAL 181  
DB 207 -----VHEVYMLRLVIRNGSVTIQNPACVYIGDKIDMKELLYPD- YQ 250  
QY 182 TLIDNRSRACH-----PCSPMKSGRCWGESEDCOSLTRVYAGGACARC--KGPL 231  
DB 251 KVTETNSHQHCYONKSKMAKCHESC-NDKCGSGDNDQRYRYSVCPSKSCQCEYSNSTS 309  
QY 232 PLDDCHECCAGCAGTPKNSDCLAHFNHSGICELCPALTYNNDFEENPNDEGRYTF 291  
DB 310 SYECCDSACLGCGTGHGRKNCIACSKYELDGLCTETCSKRIFNFKTGRVLFNPDGRION 369  
QY 292 GASCVTACPYNYL-STDVGSCTLYVC-PLHNOEYTAEDTQRCERC-SKPCARVCGLME 348  
DB 370 GNHCYKCECPPELLIENDY--CVRHCSDGHNYDATKD--VRCECKCRSSSCPKICTVDG- 423  
QY 349 HLREVRAYTSANIDCFACKCKIIFGSLAFLPESFPDQDASNTAPLQPELOVFEFLEETG 408  
DB 424 HL-----TNETLKLKECCQIDGAL-IIEHAF-----TYEDKLKLETVKVIASE 465  
QY 409 YLYISAMPDLSPLDSVFNQLOVIRGLIHANGAVSLTL-OGGLISWLGIRSLRSGSLAL 467  
DB 466 YITIV--QNNYVDLKLKLNLDIIIEGRKLIHNVKMAIATIQCCDDLEELSNLSKLIKTGAVL 523  
QY 468 IHHNTILCFVHTVPWDLF-----RNPQALLHTANRPEDCEVSGGLACHQIACARGHCWP 523  
DB 524 IMKNRRLCYVSKIDWSSIIYTSKGNKRPSLAIEARSDKLCLETEDQRYVDCMKNRKCGWK 583  
QY 524 GPYQVNCVSQFLRQGECEYECRCVLDGLPREVYNNARHCLPCHPEQOPQMSGYTCGRPADQ 583  
DB 584 PREDLCKETWKVSGTCVEKCDT-KGLRLNQTSMK-CERCSPCE-----TCNGLELD 635  
QY 584 CVACAH-----YKDPFCVACRPSGVKPDLSYMPIMKFPDEBAGACPPIINTHSCVD 636  
DB 636 CLTCHKTKLYNSDFGNRRKCVHDC-----PVSHFTQKNVCEKCHPTCY----- 679  
QY 637 LDDKCPRAEQRAPLTSONEDLG 659  
DB 680 --DNCG-----TGPDNSNG 691

Search completed: January 13, 2003, 14:49:20  
Job time : 23.4617 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 13, 2003, 14:47:00 : Search time 10.7057 Seconds  
(without alignments)  
1665.428 Million cell updates/sec

Title: US-09-854-356-6

Perfect score: 5078

Sequence: 1 METALCRMGILLALLPPGA.....TFKGTPTAENPEYGLDVPV 919

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 118974 seqs, 19401057 residues

Total number of hits satisfying chosen parameters: 118974

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Maximum Match 0%

Listing first 45 summaries

Database :

Published Applications AA:\*

- 1: /cgn2\_6/ptodata/1/pubpa/US08\_NEW\_PUB\_PEP:\*
- 2: /cgn2\_6/ptodata/1/pubpa/PCT\_NEW\_PUB\_PEP:\*
- 3: /cgn2\_6/ptodata/1/pubpa/US06\_NEW\_PUB\_PEP:\*
- 4: /cgn2\_6/ptodata/1/pubpa/US06\_PUBCOMB\_PEP:\*
- 5: /cgn2\_6/ptodata/1/pubpa/US07\_NEW\_PUB\_PEP:\*
- 6: /cgn2\_6/ptodata/1/pubpa/US07\_PUBCOMB\_PEP:\*
- 7: /cgn2\_6/ptodata/1/pubpa/PCTUS\_PUBCOMB\_PEP:\*
- 8: /cgn2\_6/ptodata/1/pubpa/US08\_PUBCOMB\_PEP:\*
- 9: /cgn2\_6/ptodata/1/pubpa/US09\_NEW\_PUB\_PEP:\*
- 10: /cgn2\_6/ptodata/1/pubpa/US09\_PUBCOMB\_PEP:\*
- 11: /cgn2\_6/ptodata/1/pubpa/US10\_NEW\_PUB\_PEP:\*
- 12: /cgn2\_6/ptodata/1/pubpa/US10\_PUBCOMB\_PEP:\*
- 13: /cgn2\_6/ptodata/1/pubpa/US60\_NEW\_PUB\_PEP:\*
- 14: /cgn2\_6/ptodata/1/pubpa/US60\_PUBCOMB\_PEP:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5078	100.0	919	9 US-09-854-356-6	Sequence 6, Appl
2	4900	96.5	1255	9 US-09-854-356-1	Sequence 1, Appl
3	4900	96.5	1255	9 US-09-930-125-2	Sequence 2, Appl
4	4892	96.3	1255	10 US-09-811-123-9	Sequence 9, Appl
5	4892	96.3	1255	10 US-09-811-115-3	Sequence 3, Appl
6	4843	95.4	1255	9 US-09-769-508-2	Sequence 2, Appl
7	4138.5	81.5	1256	9 US-09-854-356-2	Sequence 2, Appl
8	4134	81.4	1260	9 US-09-870-759-118	Sequence 118, App
9	4125.5	81.2	1256	9 US-09-854-356-14	Sequence 14, Appl
10	3954	77.9	712	9 US-09-854-356-7	Sequence 9, Appl
11	3628	71.4	653	9 US-09-854-356-3	Sequence 3, Appl
12	3590	70.7	645	10 US-09-921-161-1	Sequence 1, Appl
13	3110.5	61.3	654	9 US-09-854-356-8	Sequence 8, Appl
14	2585	50.9	564	10 US-09-821-883-3	Sequence 3, Appl
15	2585	50.9	697	10 US-09-821-883-4	Sequence 4, Appl
16	2573.5	50.7	555	10 US-09-821-883-1	Sequence 1, Appl
17	2573.5	50.7	690	10 US-09-821-883-2	Sequence 2, Appl
18	1645	33.4	1210	10 US-09-725-433-2	Sequence 2, Appl
19	1608	31.7	479	10 US-09-821-883-5	Sequence 5, Appl

20	1587	31.3	289	10 US-09-821-883-23	Sequence 23, Appl
21	1528.5	30.1	1308	10 US-09-940-101-2	Sequence 2, Appl
22	1474	29.0	583	9 US-09-930-125-9	Sequence 9, Appl
23	1474	29.0	587	9 US-09-930-125-8	Sequence 8, Appl
24	1474	29.0	589	9 US-09-930-125-10	Sequence 10, Appl
25	1474	29.0	600	9 US-09-930-125-11	Sequence 11, Appl
26	1450	28.6	266	9 US-09-854-356-4	Sequence 4, Appl
27	1447	28.5	615	10 US-09-940-101-4	Sequence 4, Appl
28	1177	23.2	217	10 US-09-821-883-25	Sequence 25, Appl
29	1177	23.2	397	10 US-09-821-883-27	Sequence 27, Appl
30	1177	23.2	1179	10 US-09-821-883-29	Sequence 29, Appl
31	1028.5	20.3	1478	10 US-09-867-521-2	Sequence 2, Appl
32	647.5	12.8	35	9 US-09-925-301-1232	Sequence 1232, Ap
33	326	6.4	59	9 US-09-854-356-5	Sequence 5, Appl
34	311.5	6.1	1367	9 US-09-870-759-120	Sequence 120, App
35	291	5.7	370	10 US-09-205-658-104	Sequence 104, App
36	291	5.7	370	10 US-09-844-353A-104	Sequence 104, App
37	267.5	5.3	1724	10 US-09-205-658-12	Sequence 12, Appl
38	267.5	5.3	1724	10 US-09-844-353A-12	Sequence 12, Appl
39	250	4.9	366	10 US-09-205-658-103	Sequence 103, App
40	250	4.9	366	10 US-09-844-353A-103	Sequence 103, App
41	223.5	4.4	383	10 US-09-205-658-105	Sequence 105, App
42	223.5	4.4	383	10 US-09-844-353A-105	Sequence 105, App
43	212	4.2	381	10 US-09-205-658-106	Sequence 106, App
44	212	4.2	381	10 US-09-844-353A-106	Sequence 106, App
45	199	3.9	830	9 US-09-870-759-140	Sequence 140, App

## ALIGNMENTS

RESULT 1  
US-09-854-356-6  
Sequence 6, Application US/09854356  
Patent No. US20020177567A1  
GENERAL INFORMATION:  
APPLICANT: Cheever, Martin A.  
APPLICANT: Cheysen, Dirk  
APPLICANT: Corixa Corporation  
TITLE OF INVENTION: HER-2/neu Fusion Proteins  
FILE REFERENCE: 014058-009810PC  
CURRENT APPLICATION NUMBER: US/09/854,356  
CURRENT FILING DATE: 2001-05-09  
PRIOR APPLICATION NUMBER: US 09/493,480  
PRIOR FILING DATE: 2000-01-28  
PRIOR APPLICATION NUMBER: US 60/117,976  
PRIOR FILING DATE: 1999-01-29  
NUMBER OF SEQ ID NOS: 26  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 6  
LENGTH: 919  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: fusion protein  
OTHER INFORMATION: Of ECD and PD of human HER-2/neu  
US-09-854-356-6  
Query Match  
Best local similarity 100.0%; Score 5078; DB 9; Length 919;  
Matches 919; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 METALCRMGILLALLPPGAASVCTGDMKRLRPASPTHMDMLRHLYGCGVQVGM 60  
Db 1 METALCRMGILLALLPPGAASVCTGDMKRLRPASPTHMDMLRHLYGCGVQVGM 60  
QY 61 ELTYLPTNASLSLQDIQYQGVVLAHQVQVPLQRLRIYNGTGLFEDNVALAVLDNG 120  
Db 61 ELTYLPTNASLSLQDIQYQGVVLAHQVQVPLQRLRIYNGTGLFEDNVALAVLDNG 120  
QY 121 DPLNNTPTVGSAGSLREIQLRSLEIFLKGVGLIQRNPQLCQDITLKKDFHKNNOA 180  
Db 121 DPLNNTPTVGSAGSLREIQLRSLEIFLKGVGLIQRNPQLCQDITLKKDFHKNNOA 180

Db 121 DPLNNTPTVTGASPGELRELOLRSTLTKGVLIIQRNPOLCYODTILMKDIFHKNNOLA 180  
Qy 181 LTLIDTNRSRACHPCSPCKSRCKGSESSDQSLTRTVCGAGCARCKGPLPTDCCHEOC 240  
Db 181 LTLIDTNRSRACHPCSPCKSRCKGSESSDQSLTRTVCGAGCARCKGPLPTDCCHEOC 240  
Qy 241 AAGCTGPKHSDCLACLFHNHSGICELHCPALVYNTDTFESMPNDEGRYTTGASCVTACP 300  
Db 241 AAGCTGPKHSDCLACLFHNHSGICELHCPALVYNTDTFESMPNDEGRYTTGASCVTACP 300  
Qy 301 YNYLSTDVGSCTLVCPRLNNOEYTAEDGTQRCCKSKPCARVCYIGMHLREBRAVTSAN 360  
Db 301 YNYLSTDVGSCTLVCPRLNNOEYTAEDGTQRCCKSKPCARVCYIGMHLREBRAVTSAN 360  
Qy 361 IQEAGCKKIRGSLAFLEPESFDGDPASNTAPLOPEQLOVFETLEITGYLYISAMPDSL 420  
Db 361 IQEAGCKKIRGSLAFLEPESFDGDPASNTAPLOPEQLOVFETLEITGYLYISAMPDSL 420  
Qy 421 DLSVFQNLQVIRGRILHNGAVSLTLOGIGISMLGRSLRELSGSLALIHNTLHCFVHTV 480  
Db 421 DLSVFQNLQVIRGRILHNGAVSLTLOGIGISMLGRSLRELSGSLALIHNTLHCFVHTV 480  
Qy 481 PMDOLFRNPQALLHTANRPEDECYEGELACHOLCARHCHGPGPTQCVNCSQFLRGQEC 540  
Db 481 PMDOLFRNPQALLHTANRPEDECYEGELACHOLCARHCHGPGPTQCVNCSQFLRGQEC 540  
Qy 541 VEECRVLOGLPREYVNAHCHLPCHECOPONGSVTCFPEADQCVACAHYKDPPECVARC 600  
Db 541 VEECRVLOGLPREYVNAHCHLPCHECOPONGSVTCFPEADQCVACAHYKDPPECVARC 600  
Qy 601 PSQVAPDLSTYMPIMKFPDEBEGACQPCPINCTHSCVDLDKCPABORASPLTSQNEIDGP 660  
Db 601 PSQVAPDLSTYMPIMKFPDEBEGACQPCPINCTHSCVDLDKCPABORASPLTSQNEIDGP 660  
Qy 661 ASPLDSTFYRSLDEDDMDGLVDAEYLYVPOGFFCPCPAPAGAGGMVHRRHSSSTRSGG 720  
Db 661 ASPLDSTFYRSLDEDDMDGLVDAEYLYVPOGFFCPCPAPAGAGGMVHRRHSSSTRSGG 720  
Qy 721 GDLTLGLEPSESEEARSPSLABSEAGSDVFDGDLGMAKGLQSLPTHTDPSPLQRYSDP 780  
Db 721 GDLTLGLEPSESEEARSPSLABSEAGSDVFDGDLGMAKGLQSLPTHTDPSPLQRYSDP 780  
Qy 781 TVPLPSELDGYVAPLTCSPQPEYVNOQPVPRQPPSPREGPLPARAPACATLERPXTLSBG 840  
Db 781 TVPLPSELDGYVAPLTCSPQPEYVNOQPVPRQPPSPREGPLPARAPACATLERPXTLSBG 840  
Qy 841 KNGVAVDVFARFAGAVENNEYLTPOGGAAPORHPAPASPADNLYYMPDODPERGAPST 900  
Db 841 KNGVAVDVFARFAGAVENNEYLTPOGGAAPORHPAPASPADNLYYMPDODPERGAPST 900  
Qy 901 FKGTPTAENPEYLGIDVPV 919  
Db 901 FKGTPTAENPEYLGIDVPV 919

RESULT 2  
US-09-854-356-1

; Sequence 1, Application US/09854356  
; Patent No. US2002017567A1  
; GENERAL INFORMATION:  
; APPLICANT: Cheever, Martin A.  
; APPLICANT: Gheysen, Dirk  
; APPLICANT: Corixa Corporation  
; APPLICANT: SmithKline Beecham Biologicals S. A.  
; TITLE OF INVENTION: HBR-2/neu Fusion Proteins  
; FILE REFERENCE: 014058-009810PC  
; CURRENT APPLICATION NUMBER: US/09/854,356  
; CURRENT FILING DATE: 2001-05-09  
; PRIOR APPLICATION NUMBER: US 09/493,480  
; PRIOR FILING DATE: 2000-01-28  
; PRIOR APPLICATION NUMBER: US 60/117,976  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 26

; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 1255  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: human HBR-2/neu protein  
; NAME/KEY: DOMAIN  
; LOCATION: (1)..(653)  
; OTHER INFORMATION: extracellular domain (ECD)  
; NAME/KEY: DOMAIN  
; LOCATION: (676)..(1255)  
; OTHER INFORMATION: intracellular domain (ICD)  
; NAME/KEY: DOMAIN  
; LOCATION: (990)..(1255)  
; OTHER INFORMATION: phosphorylation domain (PD)  
; NAME/KEY: DOMAIN  
; LOCATION: (90)..(1048)  
; OTHER INFORMATION: fragment of the phosphorylation domain, preferred  
; OTHER INFORMATION: portion (delta PD)  
US-09-854-356-1

Query Match 96.5%; Score 4900; DB 9; Length 1255;  
Best Local Similarity 73.2%; Pred. No. 5.3e-275;  
Matches 919; Conservative 0; Mismatches 0; Indels 336; Gaps 1;

Qy 1 MELALCFKMGILLALLPFGAASVQCTGDMKRLPASPETHLDMRLHYOGCQVQGNL 60  
Db 1 MELALCFKMGILLALLPFGAASVQCTGDMKRLPASPETHLDMRLHYOGCQVQGNL 60  
Qy 61 ELTYLPTNASSLFQDDIOEVQGYVLIHNOYRQVPLQRLIRVGTQLEFEDNYALAVDNG 120  
Db 61 ELTYLPTNASSLFQDDIOEVQGYVLIHNOYRQVPLQRLIRVGTQLEFEDNYALAVDNG 120  
Qy 121 DPLNNTPTVTGASPGELRELOLRSTLTKGVLIIQRNPOLCYODTILMKDIFHKNNOLA 180  
Db 121 DPLNNTPTVTGASPGELRELOLRSTLTKGVLIIQRNPOLCYODTILMKDIFHKNNOLA 180  
Qy 181 LTLIDTNRSRACHPCSPCKSRCKGSESSDQSLTRTVCGAGCARCKGPLPTDCCHEOC 240  
Db 181 LTLIDTNRSRACHPCSPCKSRCKGSESSDQSLTRTVCGAGCARCKGPLPTDCCHEOC 240  
Qy 241 AAGCTGPKHSDCLACLFHNHSGICELHCPALVYNTDTFESMPNDEGRYTTGASCVTACP 300  
Db 241 AAGCTGPKHSDCLACLFHNHSGICELHCPALVYNTDTFESMPNDEGRYTTGASCVTACP 300  
Qy 301 YNYLSTDVGSCTLVCPRLNNOEYTAEDGTQRCCKSKPCARVCYIGMHLREBRAVTSAN 360  
Db 301 YNYLSTDVGSCTLVCPRLNNOEYTAEDGTQRCCKSKPCARVCYIGMHLREBRAVTSAN 360  
Qy 361 IQEAGCKKIRGSLAFLEPESFDGDPASNTAPLOPEQLOVFETLEITGYLYISAMPDSL 420  
Db 361 IQEAGCKKIRGSLAFLEPESFDGDPASNTAPLOPEQLOVFETLEITGYLYISAMPDSL 420  
Qy 421 DLSVFQNLQVIRGRILHNGAVSLTLOGIGISMLGRSLRELSGSLALIHNTLHCFVHTV 480  
Db 421 DLSVFQNLQVIRGRILHNGAVSLTLOGIGISMLGRSLRELSGSLALIHNTLHCFVHTV 480  
Qy 481 PMDOLFRNPQALLHTANRPEDECYEGELACHOLCARHCHGPGPTQCVNCSQFLRGQEC 540  
Db 481 PMDOLFRNPQALLHTANRPEDECYEGELACHOLCARHCHGPGPTQCVNCSQFLRGQEC 540  
Qy 541 VEECRVLOGLPREYVNAHCHLPCHECOPONGSVTCFPEADQCVACAHYKDPPECVARC 600  
Db 541 VEECRVLOGLPREYVNAHCHLPCHECOPONGSVTCFPEADQCVACAHYKDPPECVARC 600  
Qy 601 PSQVAPDLSTYMPIMKFPDEBEGACQPCPINCTHSCVDLDKCPABORASPLTSQNEIDGP 660  
Db 601 PSQVAPDLSTYMPIMKFPDEBEGACQPCPINCTHSCVDLDKCPABORASPLTSQNEIDGP 660  
Qy 654 ----- 653  
Db 661 ILLVVVLGVFGLIKRQOKIRKTYMRRLLQETELVEPLTPSGAMPNOQMRILKETEL 720



QY	654	-----		653
Db	721	RKVKVLGSGAFGVYKGIWIPDGENVKIRPAVKLVIRENTSPKANKEIIDEAAYMAGVGP		780
QY	654	-----		653
Db	781	YVSRLLGICLTSTVOLVTOIMPYGCLLDHVRNCRGLSGODLLNMCQOIAKMSYLEDR		840
QY	654	-----		653
Db	841	LVHDLAARVLYKSPNHVKITDFGLARLLDDETYEHAHDSKVFIKMALESILRRFT		900
QY	654	-----		653
Db	901	HQSDVMSGYTVWELMTFGAKPYDGIIPAREIDPLEKEGRLPQPICTIDYMIWIKCM		960
QY	654	-----		653
Db	961	IDSECRPRFRELVSFSRMAKDPQRFVYUIONEDLGPASPLDSTFYRSLLIEDDDMDGLDA		1020
QY	654	-----QNEDLGPASPLDSTFYRSLLIEDDDMDGLDA		684
Db	1021	EETVLPQGGFCDDPAPGAGGWNHNRSSSTRSGGDTLTGLESEEBAPSRPLAPSEG		744
QY	745	AGSDVFEQDGLGMAAAGLSLPTNDRSPLOKRSSEDPVLPSETGUYAARLTCSPQRYV		804
Db	1081	AGSVFERGDLGMAAKGLQSLPHNDSPLOKRSSEDPVLPSETGUYAARLTCSPQRYV		1140
QY	805	NQDPVRQPPSPREGRLPARAPAGATLERKTLSPGKNGVYKDVAFGAGAVENPEYLTQ		864
Db	1141	NQDPVRQPPSPREGRLPARAPAGATLERKTLSPGKNGVYKDVAFGAGAVENPEYLTQ		1200
QY	865	GGAPRQHPRRPASPAFDNLUYDDOPRERGAPRSTFKCTPIAENPEYLTGLDVPY		919
Db	1201	GGAPRQHPRRPASPAFDNLUYDDOPRERGAPRSTFKCTPIAENPEYLTGLDVPY		1255

### RESULT 3

US-930-125-2  
SEQUENCE 2, Application US/09930125  
Application NO. US20020193329A1  
GENERAL INFORMATION:  
APPLICANT: Hand-Zimmerman, Susan  
APPLICANT: Cheever, Martin A.  
APPLICANT: Foy, Teresa M.  
APPLICANT: Lodes, Michael J.  
APPLICANT: Kalos, Michael D.  
APPLICANT: McNeill, Patricia D.  
APPLICANT: Vedvick, Thomas S.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND DIAGNOSIS  
OF HER-2/NEU-ASSOCIATED MALIGNANCIES  
FILE REFERENCE: 210121.544  
CURRENT APPLICATION NUMBER: US/09/930,125  
CURRENT FILING DATE: 2001-08-14  
NUMBER OF SEQ. ID NOS: 25  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ. ID NO. 2  
LENGTH: 1255  
TYPE: PRT  
ORGANISM: Homo sapien  
US-09-930-125-2

Query Match	96.5%	Score 4900;	DB 9;	Length 1255;
Best Local Similarity	73.2%	Pred. No. 5.3e-275;		
Matches 919; Conservative	0;	Mismatches 0;	Indels 336;	Gaps 1;

QY 1 MELALALCRWGLLALLPRGAASIVCTGTDMKRLRPA SPETHDMLRHLYQGCGVQGNL 60

Db 1 MELALALCRWGLLALLPRGAASIVCTGTDMKRLRPA SPETHDMLRHLYQGCGVQGNL 60

QY 61 ELITLPPNNAISLISFQDIQEVQGYVLIAHNOVQVPLQRLRIIVKSTQLFEDNYALAVINDG 120

Db	61	ELTILPTNASTLSIFQDIDQIEVQGVYLAHNRQVPLQRLRLYRGQVLEFDNATYALVDNG	120
Qy	121	DLPLNTPFVTGASBGBGLREJLQSLREJLKGVLQJRNPOLCYODTILMKDIFHKNNQJLA	180
Db	121	DLPLNTPFVTGASBGBGLREJLQSLREJLKGVLQJRNPOLCYODTILMKDIFHKNNQJLA	180
Qy	181	LTLIDTNSRACHCSPKCKSGSRCKWGSSESDCOSLTRTYCAGGCARCKPRLTDCCHQC	240
Db	181	LTLIDTNSRACHCSPKCKSGSRCKWGSSESDCOSLTRTYCAGGCARCKPRLTDCCHQC	240
Qy	241	AAGCGPRHSCCLCFPHNHSJGCELHCPLATYNNDFESMNPNEGRTTFEASCVTACP	300
Db	241	AAGCGPRHSCCLCFPHNHSJGCELHCPLATYNNDFESMNPNEGRTTFEASCVTACP	300
Qy	301	YNYLSTDVSGCTLYCPPLHNOEVTAAEDTORCEKCSKPCARVCYGLAMEHLREVRATVSAN	360
Db	301	YNYLSTDVSGCTLYCPPLHNOEVTAAEDTORCEKCSKPCARVCYGLAMEHLREVRATVSAN	360
Qy	361	IOEFAGCKKIFGSLAPLPESFDGDPASNTAPLQPEOLQVFEETLEETGYLTISAMPDLP	420
Db	361	IOEFAGCKKIFGSLAPLPESFDGDPASNTAPLQPEOLQVFEETLEETGYLTISAMPDLP	420
Qy	421	DLSTVQNLQVTRGKILINGAYSLTLOGLSIMJGLSLRELSGGLALHHNTHLCFVTV	480
Db	421	DLSTVQNLQVTRGKILINGAYSLTLOGLSIMJGLSLRELSGGLALHHNTHLCFVTV	480
Qy	481	PMDOLEFRPHOALHTANREDECEVGBGLACHOLCARGCMQPGTQCNCSQFLRGEC	540
Db	481	PMDOLEFRPHOALHTANREDECEVGBGLACHOLCARGCMQPGTQCNCSQFLRGEC	540
Qy	541	VEECRVLOGLPREYVNAHRCLPCHPEODPONGSVTFEGEADQCAVAANYKDPPECVARC	600
Db	541	VEECRVLOGLPREYVNAHRCLPCHPEODPONGSVTFEGEADQCAVAANYKDPPECVARC	600
Qy	601	PSGVKPDLSYMPIMKPFDEBGACQPCPINCTHSCVDLDKCGPABDRASPLTISIISAVYG	660
Db	601	PSGVKPDLSYMPIMKPFDEBGACQPCPINCTHSCVDLDKCGPABDRASPLTISIISAVYG	660
Qy	654	-----	653
Db	661	ILLYVVLGVNFGILIKRRQOKIKRYMRRLQDTELVEBELTSGAMPNOAQMRILKETEL	720
Qy	654	-----	653
Db	721	RKVVLGSGANGYTKGIMWLPDGENVKIPALIKVLEENSPRANKEIIDEAYVMAVGSP	780
Qy	654	-----	653
Db	781	YVSRLLGILSTVQVLTOLMPYGCGLDHYHRENRGRGLSQDILLNMCQJIAKMSYLEDVR	840
Qy	654	-----	653
Db	841	LVRHDLAARNLYKSPNHVKITDFGLARLLDIDETFEYHADGSKVPDKWMALESILRRFT	900
Qy	654	-----	653
Db	901	HQSDWMSGVTVWMLMFFGAKPRYDGIPIARIIDLEKGRLEPQPICTIDYUIMIMVKMM	960
Qy	654	-----	653
Db	961	IDSECRPFRELYSEFSRMAADPQRFVYQJONEDLGPASPLDSTFYRSLLIEDDDMDGLDVA	1020
Qy	654	-----	653
Db	1021	EEYLVPOQGFCCDPAGAGGMVHHRRSSSTRSGGDLTLIDLESSEEPARSPLABSEG	1080
Qy	745	AGSDVFPDGLCMGAKJLQSLPTNHDSPLOJRYSEDPVYLPSETDGYAARPLTCSQPEYV	804
Db	1081	AGSDVFPDGLCMGAKJLQSLPTNHDSPLOJRYSEDPVYLPSETDGYAARPLTCSQPEYV	1140
Qy	805	NOPIVRRQPSPREGULPAARPAAGATILERRKTLISPEKNGVAKDVFAGGAVENPEYLTPO	864
Db	1141	NOPIVRRQPSPREGULPAARPAAGATILERRKTLISPEKNGVAKDVFAGGAVENPEYLTPO	1200

QY 865 GGAAPQHPHPAPSPAFDNLXYWDODPPERGAPSTFKGTPTAENPEYLGLDVPV 919  
Db 1201 GGAAPQHPHPAPSPAFDNLXYWDODPPERGAPSTFKGTPTAENPEYLGLDVPV 1255

RESULT 4  
US-09-811-123-9  
; Sequence 9, Application US/09811123  
; Patent No. US20020001587A1  
; GENERAL INFORMATION:  
; APPLICANT: Sharon Erickson  
; APPLICANT: Ralph Schwall  
; APPLICANT: Mark Sliwowski  
; TITLE OF INVENTION: METHODS OF TREATMENT USING ANTI-ERBB  
; TITLE OF INVENTION: ANTIBODY-MAYTANSINOID CONJUGATES  
; FILE REFERENCE: GENENT 073A2  
; CURRENT APPLICATION NUMBER: US/09/811,123  
; CURRENT FILING DATE: 2001-03-16  
; PRIOR APPLICATION NUMBER: 60/238,327  
; PRIOR FILING DATE: 2000-10-05  
; PRIOR APPLICATION NUMBER: 09/602,530  
; PRIOR FILING DATE: 2000-06-23  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: FastSeq for Windows Version 4.0  
; ID NO 9  
; LENGTH: 1255  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-811-123-9

Query Match 96.3%; Score 4892; DB 10; Length 1255;  
Best Local Similarity 73.1%; Pred. No. 1,5e-274;  
Matches 918; Conservative 0; Mismatches 1; Indels 336; Gaps 1;

QY 1 METALALRWGLLALLRPGASVOTGCTDKLRLPASPERHMLMRLHYGCGVGNL 60  
Db 1 METALALRWGLLALLRPGASVOTGCTDKLRLPASPERHMLMRLHYGCGVGNL 60  
QY 61 ELTYLPTNASLFLQDIOEVGVYVLAHNOYROYPLRLIRVGTOLFEEDYALAVLDNG 120  
Db 61 ELTYLPTNASLFLQDIOEVGVYVLAHNOYROYPLRLIRVGTOLFEEDYALAVLDNG 120  
QY 121 DPLNNTPTVGTASPGGLREQLRLSLTEILKGVLIQRNPOLCYDTIIMKDIHKNNOLA 180  
Db 121 DPLNNTPTVGTASPGGLREQLRLSLTEILKGVLIQRNPOLCYDTIIMKDIHKNNOLA 180  
QY 181 LTLIDTRSRACHPCSPCKGSRGWGSSSEDCQSLTRVCAAGGACRCKGRLPTCCHEQC 240  
Db 181 LTLIDTRSRACHPCSPCKGSRGWGSSSEDCQSLTRVCAAGGACRCKGRLPTCCHEQC 240  
QY 241 AAGCTGKHSDDLACLFHNSGICELCPALVYNTDTFESMPNDEGRYTGASCYATACP 300  
Db 241 AAGCTGKHSDDLACLFHNSGICELCPALVYNTDTFESMPNDEGRYTGASCYATACP 300  
QY 301 YNLTSTVSGCTIVCPILHNOEVTAEDETQRCCKSKPCARVYCYLGMEHLREVRAYTSAN 360  
Db 301 YNLTSTVSGCTIVCPILHNOEVTAEDETQRCCKSKPCARVYCYLGMEHLREVRAYTSAN 360  
QY 361 IQERAGCKKTFGSLAFPESTGDPASNTAPLOPEQLOVFEETLEITGYLXISAMPDLP 420  
Db 361 IQERAGCKKTFGSLAFPESTGDPASNTAPLOPEQLOVFEETLEITGYLXISAMPDLP 420  
QY 421 DLSTFQMLQYIRGILNNGAVSLTLOGISLWLGRLSLRELSGIALHNTHLCEFTV 480  
Db 421 DLSTFQMLQYIRGILNNGAVSLTLOGISLWLGRLSLRELSGIALHNTHLCEFTV 480  
QY 481 PMDQLFRNPQALHTANRPEDECVGEGSLACHOLCARHCHGMPRTQCVNCSQELRGDEC 540  
Db 481 PMDQLFRNPQALHTANRPEDECVGEGSLACHOLCARHCHGMPRTQCVNCSQELRGDEC 540  
QY 541 VEECRVLQGLPREYVNAHCLPCHPECOPONGSVTCGSEPADQVACAHYDDPPCVARC 600  
Db 541 VEECRVLQGLPREYVNAHCLPCHPECOPONGSVTCGSEPADQVACAHYDDPPCVARC 600

Db 541 VEECRVLQGLPREYVNAHCLPCHPECOPONGSVTCGSEPADQVACAHYDDPPCVARC 600  
QY 601 PSYKVPDLSTYMPIMKPFDEBACOPCPIINCTHSCVDLDDKCPAEOASPLTS----- 653  
Db 601 PSYKVPDLSTYMPIMKPFDEBACOPCPIINCTHSCVDLDDKCPAEOASPLTSIVSAVVG 660  
QY 654 ----- 653  
Db 661 ILVVVLGVVFGILIKRROOKIRKTYMRLLQETELVEPLPSGAMPNOAMRLKETEL 720  
QY 654 ----- 653  
Db 721 RKVKVLGSGAFGYKGIWIPDGENVKIPVAIKVLRNTSPKANKELIDBAIYVAGVGP 780  
QY 654 ----- 653  
Db 781 YVSRLLGICLSTVQLVQLMPYGCCLLDHYRENRGLSGDOLLNMCQIAKMSYLEDVR 840  
QY 654 ----- 653  
Db 841 LVHRDLARNVLYKSPNHVKITDGLARLLDIDETEXHADGKVPKIMMALESILRRFT 900  
QY 654 ----- 653  
Db 901 HQSDVMSGYVYVMEIMTFGAKPYDGIPTAREIPDLLEKGERLPORPCTIDVYIMVCKM 960  
QY 654 -----ONEDIGPASPLDSTFYRSLSLEDDMDGLVDA 664  
Db 961 IDSECRPRRELIVSEFSRMAKDQRFVYIIONEDIGPASPLDSTFYRSLSLEDDMDGLVDA 1020  
QY 685 EBYVLPQGGFCPPRAGAGMVAHHRSSSTRSGGDLTLGLEPSEEARPSPLASEG 744  
Db 1021 EBYVLPQGGFCPPRAGAGMVAHHRSSSTRSGGDLTLGLEPSEEARPSPLASEG 1080  
QY 745 AGSDVPDGLGMAKAGLSLPTHDPSPLQRYSEDPVPLPSETDGYVAPLTCSPQPEYV 804  
Db 1081 AGSDVPDGLGMAKAGLSLPTHDPSPLQRYSEDPVPLPSETDGYVAPLTCSPQPEYV 1140  
QY 805 NOPDVRRQPSPREGRLPAARPAAGATLERPKTISPGKNGVYKDVAFAGAVENPEYLTPO 864  
Db 1141 NOPDVRRQPSPREGRLPAARPAAGATLERPKTISPGKNGVYKDVAFAGAVENPEYLTPO 1200  
QY 865 GGAAPQHPHPAPSPAFDNLXYWDODPPERGAPSTFKGTPTAENPEYLGLDVPV 919  
Db 1201 GGAAPQHPHPAPSPAFDNLXYWDODPPERGAPSTFKGTPTAENPEYLGLDVPV 1255

RESULT 5  
US-09-811-115-3  
; Sequence 3, Application US/09811115  
; Patent No. US20020035736A1  
; GENERAL INFORMATION:  
; APPLICANT: Erickson, Sharon  
; APPLICANT: Schwall, Ralph  
; APPLICANT: King, Kathleen  
; TITLE OF INVENTION: HER-2 TRANSGENIC NON-HUMAN TUMOR MODEL  
; FILE REFERENCE: GENENT 034A  
; CURRENT APPLICATION NUMBER: US/09/811,115  
; CURRENT FILING DATE: 2001-03-16  
; PRIOR APPLICATION NUMBER: 60/189,844  
; PRIOR FILING DATE: 2000-03-16  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 1255  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-811-115-3

Query Match 96.3%; Score 4892; DB 10; Length 1255;  
Best Local Similarity 73.1%; Pred. No. 1,5e-274;  
Matches 918; Conservative 0; Mismatches 1; Indels 336; Gaps 1;

QY 1 MELALACRMGLLLALLPPGAASVOVCTGTDMLRLPASPTHMLRLHLYOGQVOVGNL 60  
Db 1 MELALACRMGLLLALLPPGAASVOVCTGTDMLRLPASPTHMLRLHLYOGQVOVGNL 60  
QY 61 ELTYLPTNASTLSFLQDIOEVQGYVLLAHNOVROYPLQRLRIVRGTOLEFEDNYALAVLNG 120  
Db 61 ELTYLPTNASTLSFLQDIOEVQGYVLLAHNOVROYPLQRLRIVRGTOLEFEDNYALAVLNG 120  
QY 121 DPLNNTPTVYGASPGGRLRELOLRSLEILKGVLIOQRNPOLCYODTILMKDIFHKNNOLA 180  
Db 121 DPLNNTPTVYGASPGGRLRELOLRSLEILKGVLIOQRNPOLCYODTILMKDIFHKNNOLA 180  
QY 181 LTLIDTNRSRACHPCSPMCKGSRGWSSESDCOSLRTVYAGGACARCKGRLPTDCCHQC 240  
Db 181 LTLIDTNRSRACHPCSPMCKGSRGWSSESDCOSLRTVYAGGACARCKGRLPTDCCHQC 240  
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPREGRYTFGASCYTACP 300  
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPREGRYTFGASCYTACP 300  
QY 301 YNLTSTDVSGCTLYVCPPLHNOEVTAEADGTORCEKSKPCARVCYGLGMEHLREVRATYSAN 360  
Db 301 YNLTSTDVSGCTLYVCPPLHNOEVTAEADGTORCEKSKPCARVCYGLGMEHLREVRATYSAN 360  
QY 361 IOEPAGCKITFGSLAFIPESFDGDPASNTAPLOPEOLQVETLEETIGYLYISAMPDSL 420  
Db 361 IOEPAGCKITFGSLAFIPESFDGDPASNTAPLOPEOLQVETLEETIGYLYISAMPDSL 420  
QY 421 DLSVFOMLQYIRGRILHNGAYSLTLOGLSISWGLSRLSRLSGSLALIHNTHLCEVHTV 480  
Db 421 DLSVFOMLQYIRGRILHNGAYSLTLOGLSISWGLSRLSRLSGSLALIHNTHLCEVHTV 480  
QY 481 PMDOLFENPQALHTANRPEDDECVEGELACHOLCARGHCGWGPOTCVNCSOFLRGQEC 540  
Db 481 PMDOLFENPQALHTANRPEDDECVEGELACHOLCARGHCGWGPOTCVNCSOFLRGQEC 540  
QY 541 VEEBRVLOGLPREVYNARHCLPCHECOPONGSVTCGFPADOCVAAHKKDPFCVARC 600  
Db 541 VEEBRVLOGLPREVYNARHCLPCHECOPONGSVTCGFPADOCVAAHKKDPFCVARC 600  
QY 601 PSQVPRDLSYMPIMKFPDEGACOPCPINCTHSCVDLDDKGCAPAEORASPLTS 653  
Db 601 PSQVPRDLSYMPIMKFPDEGACOPCPINCTHSCVDLDDKGCAPAEORASPLTS 653  
QY 654 654 653  
Db 654 654 653  
QY 781 YVSLGIGLCTSTVQLVLTQMLPYGCLLDHYRENNRGRLSQDILNMCWQIAKMSYLEDDR 840  
Db 781 YVSLGIGLCTSTVQLVLTQMLPYGCLLDHYRENNRGRLSQDILNMCWQIAKMSYLEDDR 840  
QY 841 LVHRLDAAARNVLYSPNHYKITDGLARLIDIDETEXHADGKVPKIMMALESILRRRT 900  
Db 841 LVHRLDAAARNVLYSPNHYKITDGLARLIDIDETEXHADGKVPKIMMALESILRRRT 900  
QY 901 HOSDWSYGVTVWELMTFGAKPYDGLPAREIPDLLEKGEHLPOPRICTIDVYIMAKCM 960  
Db 901 HOSDWSYGVTVWELMTFGAKPYDGLPAREIPDLLEKGEHLPOPRICTIDVYIMAKCM 960  
QY 961 IDSECRPRFRELIVSEFSRMAORDPORFVYIQNEDIGPASPLDSTFYRSLLEDDDMGDLVDA 1020  
Db 961 IDSECRPRFRELIVSEFSRMAORDPORFVYIQNEDIGPASPLDSTFYRSLLEDDDMGDLVDA 1020  
QY 1021 EETLVPOQGFCCPDPAAGAGMHHRRSSSTSGGDLTLGLEPSEBEAPRSLASEG 1080  
Db 1021 EETLVPOQGFCCPDPAAGAGMHHRRSSSTSGGDLTLGLEPSEBEAPRSLASEG 1080  
QY 745 AGSDVFDGDLGMAKAGLSLPTNHDSPLORYSEDPVPLPSETDGVAVLTCSPQGEYV 804

Db 1081 AGSDVFDGDLGMAKAGLSLPTNHDSPLORYSEDPVPLPSETDGVAVLTCSPQGEYV 1140  
QY 805 NQDVRPQPPSPREGPLPAARPGATLERPKTILSPGKNVYKDVFAFGAVENPEYLTPO 864  
Db 1141 NQDVRPQPPSPREGPLPAARPGATLERPKTILSPGKNVYKDVFAFGAVENPEYLTPO 1200  
QY 865 GGAAPQPPPPASPAEDNLTYMDQDPPREGAPSPSTFKGPTAENPEYLGDLDPV 919  
Db 1201 GGAAPQPPPPASPAEDNLTYMDQDPPREGAPSPSTFKGPTAENPEYLGDLDPV 1255

RESULT 6  
US-09-769-508-2  
; Sequence 2, Application US/09769508  
; Patent No. US2002015527A1  
; GENERAL INFORMATION:  
; APPLICANT: STUART, SUSAN G.  
; APPLICANT: MONAHAN, JOHN J.  
; APPLICANT: LANGTON, BEATRICE CLAUDIA  
; APPLICANT: HANCOCK, MIRIAM E.C.  
; APPLICANT: CHAO, LORRINE A.  
; APPLICANT: BLUFORD, PETER  
; TITLE OF INVENTION: C-ERBB-2 EXTERNAL DOMAIN: GP75  
; FILE REFERENCE: BBIO-111-C1  
; CURRENT APPLICATION NUMBER: US/09/769, 508  
; NUMBER OF SEQ ID NOS: 2  
; NUMBER OF FILING DATE: 2001-01-26  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 1255  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-769-508-2

Query Match 95.4%; Score 4843; DB 9; Length 1255;  
Best Local Similarity 72.7%; Pred. No. 1e-271; 6; Indels 336; Gaps 1;  
Matches 913; Conservative 0; Mismatches

QY 1 MELALACRMGLLLALLPPGAASVOVCTGTDMLRLPASPTHMLRLHLYOGQVOVGNL 60  
Db 1 MELALACRMGLLLALLPPGAASVOVCTGTDMLRLPASPTHMLRLHLYOGQVOVGNL 60  
QY 61 ELTYLPTNASTLSFLQDIOEVQGYVLLAHNOVROYPLQRLRIVRGTOLEFEDNYALAVLNG 120  
Db 61 ELTYLPTNASTLSFLQDIOEVQGYVLLAHNOVROYPLQRLRIVRGTOLEFEDNYALAVLNG 120  
QY 121 DPLNNTPTVYGASPGGRLRELOLRSLEILKGVLIOQRNPOLCYODTILMKDIFHKNNOLA 180  
Db 121 DPLNNTPTVYGASPGGRLRELOLRSLEILKGVLIOQRNPOLCYODTILMKDIFHKNNOLA 180  
QY 181 LTLIDTNRSRACHPCSPMCKGSRGWSSESDCOSLRTVYAGGACARCKGRLPTDCCHQC 240  
Db 181 LTLIDTNRSRACHPCSPMCKGSRGWSSESDCOSLRTVYAGGACARCKGRLPTDCCHQC 240  
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPREGRYTFGASCYTACP 300  
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPREGRYTFGASCYTACP 300  
QY 301 YNLTSTDVSGCTLYVCPPLHNOEVTAEADGTORCEKSKPCARVCYGLGMEHLREVRATYSAN 360  
Db 301 YNLTSTDVSGCTLYVCPPLHNOEVTAEADGTORCEKSKPCARVCYGLGMEHLREVRATYSAN 360  
QY 361 IOEPAGCKITFGSLAFIPESFDGDPASNTAPLOPEOLQVETLEETIGYLYISAMPDSL 420  
Db 361 IOEPAGCKITFGSLAFIPESFDGDPASNTAPLOPEOLQVETLEETIGYLYISAMPDSL 420  
QY 421 DLSVFOMLQYIRGRILHNGAYSLTLOGLSISWGLSRLSRLSGSLALIHNTHLCEVHTV 480  
Db 421 DLSVFOMLQYIRGRILHNGAYSLTLOGLSISWGLSRLSRLSGSLALIHNTHLCEVHTV 480  
QY 481 PMDOLFENPQALHTANRPEDDECVEGELACHOLCARGHCGWGPOTCVNCSOFLRGQEC 540  
Db 481 PMDOLFENPQALHTANRPEDDECVEGELACHOLCARGHCGWGPOTCVNCSOFLRGQEC 540

```
Db 481 PWDOLFRRPHOALLHTANRPEDCEYEGSLACHQOLCARALLGSGPTQCVCNSQFLRGDC 540
Qy 541 VEECRVLQGLREYVYNAHCLPCHEPCOPONGSVTCFGEPAQCVAAHYKDPPECVARC 600
Db 541 VEECRVLQGLREYVYNAHCLPCHEPCOPONGSVTCFGEPAQCVAAHYKDPPECVARC 600
Qy 601 PSYGVPDLSTYPIWKFPDEBACQPCPINCCHSCVDLDDKGCAPARQASPLTS----- 653
Db 601 PSYGVPDLSTYPIWKFPDEBACQPCPINCCHSCVDLDDKGCAPARQASPLTSIVSAVVG 660
Qy 654 ----- 653
Db 661 ILLVVLGVFGLIKRROQKIRKTYMRLLQETELVEPLTPSGAMPNQAOMRIKETEL 720
Qy 654 ----- 653
Db 721 RKKVLGSGAFYVYKGIWIDGENVKIPVAIKVLENTSPKANKELDEAVYAGVGP 780
Qy 654 ----- 653
Db 781 YVSRLLIGLSTVOLVTLQMLPYGCLLDHVRENRGLGSODLLNMCQIANKMSYLEDVR 840
Qy 654 ----- 653
Db 841 LVHRDLAARNVLYKSPNHVKITDEGLARLLIDETEHADGKVPDKMALESILRRRT 900
Qy 654 ----- 653
Db 901 HOSDVASYGTVMELMTGAKPYDGIAPAREIPDLLEKEGBRLPQPICTIDVYMIWVKCM 960
Qy 654 ----- 684
Db 961 IDSECRPRRELVSFSMARDPQRFVYIQNEDLGAPSLDSTFYRSLEDDMDGLDYDA 1020
Qy 685 EEYLPOOGFCFDPAPAGGKVVHRRSSSTRSGGDLTGLSESEBAPRSLAPSG 744
Db 1021 EEYLPOOGFCFDPAPAGGKVVHRRSSSTRSGGDLTGLSESEBAPRSLAPSG 1080
Qy 745 AGSDVFDGLMGAAKGLSLPTHDSPLOKRSSEDPVPLPSETDGYAVALTCSPOPEY 804
Db 1081 AGSDVFDGLMGAAKGLSLPTHDSPLOKRSSEDPVPLPSETDGYAVALTCSPOPEY 1140
Qy 805 NOPDVRRPSPSPRSGPLPAARPAATLERPKTSLSGKNGVYKDVAFGGAVENTPEYLPQ 864
Db 1141 NOPDVRRPSPSPRSGPLPAARPAATLERPKTSLSGKNGVYKDVAFGGAVENTPEYLPQ 1200
Qy 865 GGAAPQHPHPAPSPAFNDLYWMDPPERGAPESTFKGTPLAENPEYLGLDVPY 919
Db 1201 GGAAPQHPHPAPSPAFNDLYWMDPPERGAPESTFKGTPLAENPEYLGLDVPY 1255

RESULT 7
US-09-854-356-2
: Sequence 2, Application US/09854356
: Patent No. US20020177567A1
: GENERAL INFORMATION:
: APPLICANT: Cheever, Martin A.
: APPLICANT: Gheysen, Dirk
: APPLICANT: Corixa Corporation
: APPLICANT: SmithKline Beecham Biologicals S. A.
: TITLE OF INVENTION: HER-2/neu Fusion Proteins
: FILE REFERENCE: 014058-009810PC
: CURRENT APPLICATION NUMBER: US/09/854,356
: PRIOR FILING DATE: 2001-05-09
: PRIOR APPLICATION NUMBER: US 09/493,480
: PRIOR FILING DATE: 2000-01-28
: PRIOR APPLICATION NUMBER: US 60/117,976
: NUMBER OF SEQ ID NOS: 26
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 2
: LENGTH: 1256
: TYPE: PRT
```

```
;; ORGANISM: Rattus sp.
;;
;; FEATURE:
;; OTHER INFORMATION: rat HER-2/neu protein
;; NAME/KEY: DOMAIN
;; LOCATION: (1)..(654)
;; OTHER INFORMATION: extracellular domain (ECD)
;; NAME/KEY: DOMAIN
;; LOCATION: (677)..(1256)
;; OTHER INFORMATION: intracellular domain (ICD)
;; NAME/KEY: DOMAIN
;; LOCATION: (721)..(998)
;; OTHER INFORMATION: kinase domain (KD)
;; NAME/KEY: DOMAIN
;; LOCATION: (991)..(1256)
;; OTHER INFORMATION: phosphorylation domain (PD)
;; NAME/KEY: DOMAIN
;; LOCATION: (991)..(1049)
;; OTHER INFORMATION: fragment of the phosphorylation domain, preferred
;; OTHER INFORMATION: portion (delta PD)
US-09-854-356-2
```

Query Match 81.5%; Score 4138.5; DB 9; Length 1256;  
Best Local Similarity 62.2%; Pred. No. 4.1e-231;  
Matches 781; Conservative 42; Mismatches 96; Indels 337; Gaps 2;

```
Qy 1 MELALCRWGLLLALPPGAASVCTGTGDKRLRLPASPEHLMLRLHYGCGVYQGNL 60
Db 1 MELAMCRWGLLLALPPGIAAGVCTGTGDKRLRLPASPEHLMLRLHYGCGVYQGNL 60
Qy 61 ELTYLPTNASTSLFLODIDVQGYVLIANNQVQVPLQRLRYRGQLFEDYATLAVLDNG 120
Db 61 ELTYLPANASTSLFLODIDVQGYVLIANNQVQVPLQRLRYRGQLFEDYATLAVLDNR 120
Qy 121 DPLNNTTPVT-GASPGGLRELOLRSLTEILKGVLIQNNPOLCYODTIMKDIFFHKNQL 179
Db 121 DPLNNTTPVT-GASPGGLRELOLRSLTEILKGVLIQNNPOLCYODTIMKDIFFHKNQL 179
Qy 180 ALTLIDNRSRACHPSPKSGSRGSGSSSEDCQSLRTVACAGGACGKPLPTDCHEQ 239
Db 180 ALTLIDNRSRACHPSPKSGSRGSGSSSEDCQSLRTVACAGGACGKPLPTDCHEQ 239
Qy 240 CAAGCTGKHSCLACLAFHNSGICELHCPALVYNTDTFESMPRDEGRYTFGASCYVAC 299
Db 240 CAAGCTGKHSCLACLAFHNSGICELHCPALVYNTDTFESMPRDEGRYTFGASCYVAC 299
Qy 300 PNYLSTVSGCTIYCPHNOEVTAEQNGOREKSKCARCYGLGMEHLREYAVTSA 359
Db 300 PNYLSTVSGCTIYCPHNOEVTAEQNGOREKSKCARCYGLGMEHLREYAVTSA 359
Qy 360 NIOEPAGCKITFGSLAFIPESFDGDPASNTAPLQEQLOVEFTEITGYLYISAMPDSL 419
Db 360 NIOEPAGCKITFGSLAFIPESFDGDPASNTAPLQEQLOVEFTEITGYLYISAMPDSL 419
Qy 420 PDLSTFQNLQVYRGILNNGAYSLTLOGIGISMLRLSLRELGSALAIHHNTHLCFYHT 479
Db 420 PDLSTFQNLQVYRGILNNGAYSLTLOGIGISMLRLSLRELGSALAIHHNTHLCFYHT 479
Qy 480 VPMDOLFRRPHOALLHTANRPEDCEYEGSLACHQOLCARALLGSGPTQCVCNSQFLRGDC 539
Db 480 VPMDOLFRRPHOALLHTANRPEDCEYEGSLACHQOLCARALLGSGPTQCVCNSQFLRGDC 540
Qy 540 VEECRVLQGLREYVYNAHCLPCHEPCOPONGSVTCFGEPAQCVAAHYKDPPECVAR 599
Db 540 VEECRVLQGLREYVYNAHCLPCHEPCOPONGSVTCFGEPAQCVAAHYKDPPECVAR 600
Qy 600 CPSGVKPDLSYPIWKFPDEBACQPCPINCCHSCVDLDDKGCAPARQASPLT----- 652
Db 600 CPSGVKPDLSYPIWKFPDEBACQPCPINCCHSCVDLDDKGCAPARQASPLT----- 652
Qy 652 ----- 652
Db 661 GYLLELLVYVVGIIIKRRROKIRKTYMRLLQETELVEPLTPSGAMPNQAOMRIKETE 720
```

QY 653 ----- 652  
Db 721 LRKVVVLSGAGFTYKGIWIPDGENVKIPVAIKLRENTSPKANKEILDEAYVAGVS 780  
QY 653 ----- 652  
Db 781 PYVSRLGICLTSTVOLVQMLPYGCLLDHVRHGRGLSQDILNMCVQIAKMSYLEDY 840  
QY 653 ----- 652  
Db 841 RLVRDLAARNVLYKSPNHVKITDFGLARLDIDETEHADGKVPDKMMALLESILRRRF 900  
QY 653 ----- 652  
Db 901 THOSDVMSTGYVTWELMTFGAKPYDGIPIARELPDLLEKGERLPDPICITIDVYIMVCKM 960  
QY 653 ----- 652  
Db 961 MIDSECRPRFRELVESEFARMARDPQRFVVIQNEDELGPSSPMOSTFYRSLLEDMDMDGLVD 1020  
QY 684 ABEYLVPOQGFPCPDPAAGAGVHHRRSSSTRSGGDLTLGLEPSEEDAPRSLAPSE 743  
Db 1021 ABEYLVPOQGFSPDPPTGTAHRRRHSSTRSGGDLTLGLEPSEEGPRPSPLAPSE 1080  
QY 744 GAGSDVFDGDLGAKGLQSLPTHDPSPLORYSEDPVPLPSETPDGYVAPLTCSPOPEY 803  
Db 1081 GAGSDVFDGDLGAKGLQSLPTHDPSPLORYSEDPVPLPSETPDGYVAPLTCSPOPEY 1140  
QY 804 VNQPDVPRPQPSRREGPLPAARAGATLERPKTLSPGKNGVYKDYAFAGAVENPEYLV 863  
Db 1141 VNQSEVOQPPPLTREGPLPRVPRAGATLERPKTLSPGKNGVYKDYAFAGAVENPEYLV 1200  
QY 864 OGGAPOHPHPPAPSPAFDNLTYWDOPPEEGAPRSTFGKPTAENPEYLGADV 919  
Db 1201 REGTASPPHPPAPSPAFDNLTYWDONSSSEGGPPSPNEGPTAENPEYLGADV 1256

## RESULT 8

US-09-870-759-118  
Sequence 118, Application US/09870759  
Patent No. US20020177551A1  
GENERAL INFORMATION:  
APPLICANT: TERMAN, David S  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT OF NEOPLASTIC DISEASE  
FILE REFERENCE: 870759  
CURRENT APPLICATION NUMBER: US/09/870,759  
PRIOR FILING DATE: 2002-01-14  
PRIOR APPLICATION NUMBER: US 60/208,128  
PRIOR FILING DATE: 2000-05-30  
NUMBER OF SEQ ID NOS: 166  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 118  
LENGTH: 1260  
TYPE: PRT  
ORGANISM: Rattus norvegicus  
US-09-870-759-118

Query Match 81.4%; Score 4134; DB 9; Length 1260;  
Best Local Similarity 62.1%; Pred. No. 7,4e-231;  
Matches 781; Conservative 42; Mismatches 96; Indels 338; Gaps 3;

QY 1 METALACRMGLLALLPAGASTOVCTGTMKRLPASPETHLMDMRLHYOGGOVVOGML 60  
Db 4 MELAAACRMGFLALLPPIAGTQVCTGTMKRLPASPETHLMDMRLHYOGGOVVOGML 63  
QY 61 EETLYPTNASLSFLQDIEVOGVYLIANHOVROVPLQRLRIYRGTOLEFDNTALAVLDNG 120  
Db 64 EETLYPTNASLSFLQDIEVOGVYLIANHOVROVPLQRLRIYRGTOLEFDNTALAVLDNR 123  
QY 121 DPLANTTPVT-GASPGGLRELDRLSTETLKGCVLIQNRPOLCYODTILMKDIFRKNOL 179  
Db 124 DPQDNVAASTPGRTPEGLREQLRLSTETILKGCVLIRGNPOLCYODMVLAKDVFRRKNOL 183

QY 180 ALTLIDNRSRACHPCSPMCKSGRCWGESSEDOQSLTRIVYACAGGACRCKPIPTDCCHEO 239  
Db 184 APYDIDTNRSRACPCAPACKDHCWGESSEDOQSLTRIVYACAGGACRCKPIPTDCCHEO 243  
QY 240 CAAGCTGPKHSDIACLIHNSGICELHC PALVYNTDFFESMNPREGRTYGASCVTAC 299  
Db 244 CAAGCTGPKHSDIACLIHNSGICELHC PALVYNTDFFESMNPREGRTYGASCVTAC 303  
QY 300 PNYVLTSDVSGCTLYVCLPHNQEVTAEDGTQCEKSKPCARVCYGLGMEHLREAVTSA 359  
Db 304 PNYVLTSDVSGCTLYVCLPHNQEVTAEDGTQCEKSKPCARVCYGLGMEHLREAVTSA 363  
QY 360 NIOEPAGCKITIGSLAFLEPSEFDGDPASNTAPLQPPQLOVFELEITITGTYISANPDSL 419  
Db 364 NIOEPAGCKITIGSLAFLEPSEFDGDPASNTAPLQPPQLOVFELEITITGTYISANPDSL 423  
QY 420 PDLVSQNLQVIRGRILHNGAVSLTLQGGISWLGIRSLRELGSGALJHNTHTLCEVHT 479  
Db 424 RDLVSQNLQVIRGRILHNGAVSLTLQGGISWLGIRSLRELGSGALJHNTHTLCEVHT 483  
QY 480 VPMDOLEFRNPQALLHTANRPEDE-CVGEGLACHOLCARGHCWGPGPTQCVNCSOFLRQ 538  
Db 484 VPMDOLEFRNPQALLHTANRPEDE-CVGEGLACHOLCARGHCWGPGPTQCVNCSOFLRQ 543  
QY 539 ECVEECRVLOGLPREVYNARHCLPCHPECOPONGSVTCGPPEADOCVAAHKKDPEFCVA 598  
Db 544 ECVEECRVLOGLPREVYNARHCLPCHPECOPONGSVTCGPPEADOCVAAHKKDPEFCVA 603  
QY 599 RCPGCVKPLDLYMPIMKFPDEGACQPCP INCTHSCVDLDDKCPAEGASPLT----- 652  
Db 604 RCPGCVKPLDLYMPIMKFPDEGACQPCP INCTHSCVDLDDKCPAEGASPLT----- 663  
QY 653 ----- 652  
Db 664 EGVLLFLLVYVYGLIKRRROKIRKTYMRLLQETELVEPLTPSGAMPNQAMRLKET 723  
QY 653 ----- 652  
Db 724 ELRKVYVLSGAGFTYKGIWIPDGENVKIPVAIKLRENTSPKANKEILDEAYVAGVG 783  
QY 653 ----- 652  
Db 784 SPVVSRLGICLTSTVOLVQMLPYGCLLDHVRHGRGLSQDILNMCVQIAKMSYLEDY 843  
QY 653 ----- 652  
Db 844 VRLVRDLAARNVLYKSPNHVKITDFGLARLDIDETEHADGKVPDKMMALLESILRRR 903  
QY 653 ----- 652  
Db 904 FTHOSDVMSTGYVTWELMTFGAKPYDGIPIARELPDLLEKGERLPDPICITIDVYIMVCK 963  
QY 653 ----- 652  
Db 964 WIDSECRPRFRELVESEFARMARDPQRFVVIQNEDELGPSSPMOSTFYRSLLEDMDMDGLVD 1023  
QY 683 DAEYLVPOQGFPCPDPAAGAGVHHRRSSSTRSGGDLTLGLEPSEEDAPRSLAPSE 742  
Db 1024 DAEYLVPOQGFSPDPPTGTAHRRRHSSTRSGGDLTLGLEPSEEGPRPSPLAPSE 1083  
QY 743 EGAGSDVFDGDLGAKGLQSLPTHDPSPLORYSEDPVPLPSETPDGYVAPLTCSPOPE 802  
Db 1084 EGAGSDVFDGDLGAKGLQSLPTHDPSPLORYSEDPVPLPSETPDGYVAPLTCSPOPE 1143  
QY 803 VYNQPDVPRPQPSRREGPLPAARAGATLERPKTLSPGKNGVYKDYAFAGAVENPEYLV 862  
Db 1144 VYNQSEVOQPPPLTREGPLPRVPRAGATLERPKTLSPGKNGVYKDYAFAGAVENPEYLV 1203  
QY 863 POGGAPOHPHPPAPSPAFDNLTYWDOPPEEGAPRSTFGKPTAENPEYLGADV 919  
Db 1204 PREGTASPPHPPAPSPAFDNLTYWDONSSSEGGPPSPNEGPTAENPEYLGADV 1260



```
Matches 712; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MELAALCRWGLLALLPRGAASVQVCTGTMKRLRASPETHLDMRLHYLQGCQVQVGNL 60
DB 1 MELAALCRWGLLALLPRGAASVQVCTGTMKRLRASPETHLDMRLHYLQGCQVQVGNL 60
QY 61 ELYLPFNASLSFLQDIQEVQGVYLLAHNQVROPVLRIRYVGTQLFEDNVALAVLDNG 120
DB 61 ELYLPFNASLSFLQDIQEVQGVYLLAHNQVROPVLRIRYVGTQLFEDNVALAVLDNG 120
QY 121 DPLNNTPTVYGASPGGLRELQRLSLTEILKGVLLIORNPOLCYQDTILMKDIFHKNNQLA 180
DB 121 DPLNNTPTVYGASPGGLRELQRLSLTEILKGVLLIORNPOLCYQDTILMKDIFHKNNQLA 180
QY 181 LTLIDNRSRACHPCSPMKGSRKMGESSEDDCSLTRVACAGCARKGRLPTDCCHQEC 240
DB 181 LTLIDNRSRACHPCSPMKGSRKMGESSEDDCSLTRVACAGCARKGRLPTDCCHQEC 240
QY 241 AAGCTGPKHSDDCLACHFNHSGICEHLCPALVYNTDFESMPNPGRYTFGASCTYAC 300
DB 241 AAGCTGPKHSDDCLACHFNHSGICEHLCPALVYNTDFESMPNPGRYTFGASCTYAC 300
QY 301 YNLTSTDVGSCTLVCPHLNQEYTAEDGTQRCCKSPCARVCYGLGMEHLREVAVTSAN 360
DB 301 YNLTSTDVGSCTLVCPHLNQEYTAEDGTQRCCKSPCARVCYGLGMEHLREVAVTSAN 360
QY 361 IOEFACKKIFGSLAFPLPSFDGDPASNTAPLOPEOLOVPEETLEETGYLYISAMPDSL 420
DB 361 IOEFACKKIFGSLAFPLPSFDGDPASNTAPLOPEOLOVPEETLEETGYLYISAMPDSL 420
QY 421 DLSVFONLQVIRIRILHNGAYSLTLOGLSIWLGLSLRELGSGLALIHNNHILCFVHY 480
DB 421 DLSVFONLQVIRIRILHNGAYSLTLOGLSIWLGLSLRELGSGLALIHNNHILCFVHY 480
QY 481 PMQDLFRNHQALLHTANRPEDECVGEGLAGHQLARGHCWGPGPTQVCNCSQFLRGQEC 540
DB 481 PMQDLFRNHQALLHTANRPEDECVGEGLAGHQLARGHCWGPGPTQVCNCSQFLRGQEC 540
QY 541 VECRVLOGLPREYVNAHCLPCHPEOPONGSVTCFGEADQCVACAHYKDRPFCVARC 600
DB 541 VECRVLOGLPREYVNAHCLPCHPEOPONGSVTCFGEADQCVACAHYKDRPFCVARC 600
QY 601 PSGVKPDLSTYMPIMKRPDEEGACQPCPINCTHSCVDLDKGCRAEDRASPLTSQNDLGR 660
DB 601 PSGVKPDLSTYMPIMKRPDEEGACQPCPINCTHSCVDLDKGCRAEDRASPLTSQNDLGR 660
b 661 ASPLDSTFYRSLLEDDMGDLVDAEYLVVQGFECRDPARPGAGVNNHNR 712
b 661 ASPLDSTFYRSLLEDDMGDLVDAEYLVVQGFECRDPARPGAGVNNHNR 712

RESULT 11
US-09-854-356-3
; Sequence 3, Application US/09854356
; Patent No. US20020177567A1
; GENERAL INFORMATION:
; APPLICANT: Cheever, Martin A.
; APPLICANT: Gheysen, Dirk
; APPLICANT: Corixa Corporation
; APPLICANT: SmithKline Beecham Biologicals S. A.
; TITLE OF INVENTION: HER-2/neu Fusion Proteins
; FILE REFERENCE: 014058-009810PC
; CURRENT APPLICATION NUMBER: US/09/854,356
; CURRENT FILING DATE: 2001-05-09
; PRIOR APPLICATION NUMBER: US 09/493,480
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: US 60/117,976
; PRIOR FILING DATE: 1998-01-29
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 653
; TYPE: PRT
```

```
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: extracellular domain (ECD) of human HER-2/neu
US-09-854-356-3
Query Match
Best local similarity 100.0%; Pred. No. 5,1e-202;
Matches 653; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
71.4%; Score 3628; DB 9; Length 653;

QY 1 MELAALCRWGLLALLPRGAASVQVCTGTMKRLRASPETHLDMRLHYLQGCQVQVGNL 60
DB 1 MELAALCRWGLLALLPRGAASVQVCTGTMKRLRASPETHLDMRLHYLQGCQVQVGNL 60
QY 61 ELYLPFNASLSFLQDIQEVQGVYLLAHNQVROPVLRIRYVGTQLFEDNVALAVLDNG 120
DB 61 ELYLPFNASLSFLQDIQEVQGVYLLAHNQVROPVLRIRYVGTQLFEDNVALAVLDNG 120
QY 121 DPLNNTPTVYGASPGGLRELQRLSLTEILKGVLLIORNPOLCYQDTILMKDIFHKNNQLA 180
DB 121 DPLNNTPTVYGASPGGLRELQRLSLTEILKGVLLIORNPOLCYQDTILMKDIFHKNNQLA 180
QY 181 LTLIDNRSRACHPCSPMKGSRKMGESSEDDCSLTRVACAGCARKGRLPTDCCHQEC 240
DB 181 LTLIDNRSRACHPCSPMKGSRKMGESSEDDCSLTRVACAGCARKGRLPTDCCHQEC 240
QY 241 AAGCTGPKHSDDCLACHFNHSGICEHLCPALVYNTDFESMPNPGRYTFGASCTYAC 300
DB 241 AAGCTGPKHSDDCLACHFNHSGICEHLCPALVYNTDFESMPNPGRYTFGASCTYAC 300
QY 301 YNLTSTDVGSCTLVCPHLNQEYTAEDGTQRCCKSPCARVCYGLGMEHLREVAVTSAN 360
DB 301 YNLTSTDVGSCTLVCPHLNQEYTAEDGTQRCCKSPCARVCYGLGMEHLREVAVTSAN 360
QY 361 IOEFACKKIFGSLAFPLPSFDGDPASNTAPLOPEOLOVPEETLEETGYLYISAMPDSL 420
DB 361 IOEFACKKIFGSLAFPLPSFDGDPASNTAPLOPEOLOVPEETLEETGYLYISAMPDSL 420
QY 421 DLSVFONLQVIRIRILHNGAYSLTLOGLSIWLGLSLRELGSGLALIHNNHILCFVHY 480
DB 421 DLSVFONLQVIRIRILHNGAYSLTLOGLSIWLGLSLRELGSGLALIHNNHILCFVHY 480
QY 481 PMQDLFRNHQALLHTANRPEDECVGEGLAGHQLARGHCWGPGPTQVCNCSQFLRGQEC 540
DB 481 PMQDLFRNHQALLHTANRPEDECVGEGLAGHQLARGHCWGPGPTQVCNCSQFLRGQEC 540
QY 541 VECRVLOGLPREYVNAHCLPCHPEOPONGSVTCFGEADQCVACAHYKDRPFCVARC 600
DB 541 VECRVLOGLPREYVNAHCLPCHPEOPONGSVTCFGEADQCVACAHYKDRPFCVARC 600
QY 601 PSGVKPDLSTYMPIMKRPDEEGACQPCPINCTHSCVDLDKGCRAEDRASPLTS 653
DB 601 PSGVKPDLSTYMPIMKRPDEEGACQPCPINCTHSCVDLDKGCRAEDRASPLTS 653

RESULT 12
US-09-921-161-1
; Sequence 1, Application US/09921161
; Patent No. US2002090662A1
; GENERAL INFORMATION:
; APPLICANT: Ralph, Peter
; APPLICANT: ANALYTICAL METHOD
; TITLE OF INVENTION: ANALYTICAL METHOD
; FILE REFERENCE: GENENT 066A
; CURRENT APPLICATION NUMBER: US/09/921,161
; CURRENT FILING DATE: 2001-08-01
; PRIOR APPLICATION NUMBER: 60/225,433
; PRIOR FILING DATE: 2000-08-15
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 645
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-921-161-1
```

Query Match 70.7%; Score 3590; DB 10; Length 645;  
Best Local Similarity 100.0%; Pred. No. 7.8e-200;  
Matches 645; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 MELAAACRMGGLLALLPPGAASVGYCTGDMKRLPASPETHLMDRLHYOGCCVYQGNL 60
DB 1 MELAAACRMGGLLALLPPGAASVGYCTGDMKRLPASPETHLMDRLHYOGCCVYQGNL 60
QY 61 ELTYLPTNASSFLQDIOEVGYLIIAHNOYRVPLQRLRIYRGTLQFEDNVALAVLNG 120
DB 61 ELTYLPTNASSFLQDIOEVGYLIIAHNOYRVPLQRLRIYRGTLQFEDNVALAVLNG 120
QY 121 DPLNNTPTVTSASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILMKDIFHKNNQ 180
DB 121 DPLNNTPTVTSASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILMKDIFHKNNQ 180
QY 181 LTLIDTNRSRACHPCSPMKSGRCWGESSEDCQSLRTVCAGGACRGKPLPTDCCHEQ 240
DB 181 LTLIDTNRSRACHPCSPMKSGRCWGESSEDCQSLRTVCAGGACRGKPLPTDCCHEQ 240
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMNPDEGRYTFGASCVTAC 300
DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMNPDEGRYTFGASCVTAC 300
QY 301 YNYLSTDVGSCTLVCPILHNOEYTAEDGTQRCCKSKPCARVCYGLGMEHLREVRATVSA 360
DB 301 YNYLSTDVGSCTLVCPILHNOEYTAEDGTQRCCKSKPCARVCYGLGMEHLREVRATVSA 360
QY 361 IOEFAGCKKIFGSLAFLESPFDGDPASNTAPLOPEQLOVEFTELEITGYLIYSAMPDSL 420
DB 361 IOEFAGCKKIFGSLAFLESPFDGDPASNTAPLOPEQLOVEFTELEITGYLIYSAMPDSL 420
QY 421 DLSVFOQLVIRGRILHNGAYSLTLOGLISWLGSLRSLRELSGLALIHNTLCEFTV 480
DB 421 DLSVFOQLVIRGRILHNGAYSLTLOGLISWLGSLRSLRELSGLALIHNTLCEFTV 480
QY 481 PWDOLFRRNPQALHTANRPEDECVGEGSLACHQLCARGHCWGPPPTQCVNCSOFLRGQEC 540
DB 481 PWDOLFRRNPQALHTANRPEDECVGEGSLACHQLCARGHCWGPPPTQCVNCSOFLRGQEC 540
QY 541 VEEECVTLQGLPREYVNAHCLPCHPECOPONGSVTCFGEADQCAACAHYKDPFCVVARC 600
DB 541 VEEECVTLQGLPREYVNAHCLPCHPECOPONGSVTCFGEADQCAACAHYKDPFCVVARC 600
QY 601 PSQVPRDLSTYPIWKFPDEEGACQPCPINCTHSCVDLDERGCPAQRASPV 654
DB 601 PSQVPRDLSTYPIWKFPDEEGACQPCPINCTHSCVDLDERGCPAQRASPV 654
```

RESULT 13  
US-09-854-356-8  
Sequence 8, Application US/09854356  
Patent No. US2002017567A1  
GENERAL INFORMATION:  
APPLICANT: Cheever, Martin A.  
APPLICANT: Gheysen, Dirk  
APPLICANT: Corixa Corporation  
APPLICANT: SmithKline Beecham Biologicals S. A.  
TITLE OF INVENTION: HER-2/neu Fusion Proteins  
FILE REFERENCE: 014058-009810PC  
CURRENT APPLICATION NUMBER: US/09/854,356  
CURRENT FILING DATE: 2001-03-09  
PRIORITY APPLICATION NUMBER: US 09/493,480  
PRIORITY FILING DATE: 2000-01-28  
PRIOR APPLICATION NUMBER: US 60/117,976  
PRIOR FILING DATE: 1999-01-29  
NUMBER OF SEQ ID NOS: 26  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 8  
LENGTH: 654  
TYPE: PRT  
ORGANISM: Rattus sp.

FEATURE:  
OTHER INFORMATION: extracellular domain (ECD) of rat HER-2/neu  
US-09-854-356-8

Query Match 61.3%; Score 3110.5; DB 9; Length 654;  
Best Local Similarity 85.5%; Pred. No. 3.4e-112;  
Matches 558; Conservative 32; Mismatches 62; Indels 1; Gaps 1;

```
QY 1 MELAAACRMGGLLALLPPGAASVGYCTGDMKRLPASPETHLMDRLHYOGCCVYQGNL 60
DB 1 MELAAACRMGGLLALLPPGAASVGYCTGDMKRLPASPETHLMDRLHYOGCCVYQGNL 60
QY 61 ELTYLPTNASSFLQDIOEVGYLIIAHNOYRVPLQRLRIYRGTLQFEDNVALAVLNG 120
DB 61 ELTYLPTNASSFLQDIOEVGYLIIAHNOYRVPLQRLRIYRGTLQFEDNVALAVLNG 120
QY 121 DPLNNTPTVTSASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILMKDIFHKNNQ 179
DB 121 DPLNNTPTVTSASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILMKDIFHKNNQ 180
QY 180 ALTLIDTNRSRACHPCSPMKSGRCWGESSEDCQSLRTVCAGGACRGKPLPTDCCHEQ 239
DB 180 ALTLIDTNRSRACHPCSPMKSGRCWGESSEDCQSLRTVCAGGACRGKPLPTDCCHEQ 240
QY 240 CAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMNPDEGRYTFGASCVTAC 299
DB 240 CAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMNPDEGRYTFGASCVTAC 300
QY 300 PNYLSTDVGSCTLVCPILHNOEYTAEDGTQRCCKSKPCARVCYGLGMEHLREVRATVSA 359
DB 300 PNYLSTDVGSCTLVCPILHNOEYTAEDGTQRCCKSKPCARVCYGLGMEHLREVRATVSA 360
QY 360 NIOEFAGCKKIFGSLAFLESPFDGDPASNTAPLOPEQLOVEFTELEITGYLIYSAMPDSL 419
DB 360 NIOEFAGCKKIFGSLAFLESPFDGDPASNTAPLOPEQLOVEFTELEITGYLIYSAMPDSL 420
QY 420 PDLSTFQQLVIRGRILHNGAYSLTLOGLISWLGSLRSLRELSGLALIHNTLCEFTV 479
DB 420 PDLSTFQQLVIRGRILHNGAYSLTLOGLISWLGSLRSLRELSGLALIHNTLCEFTV 480
QY 480 VPDOLFRRNPQALHTANRPEDECVGEGSLACHQLCARGHCWGPPPTQCVNCSOFLRGQEC 539
DB 480 VPDOLFRRNPQALHTANRPEDECVGEGSLACHQLCARGHCWGPPPTQCVNCSOFLRGQEC 540
QY 540 CVEECVTLQGLPREYVNAHCLPCHPECOPONGSVTCFGEADQCAACAHYKDPFCVVAR 599
DB 540 CVEECVTLQGLPREYVNAHCLPCHPECOPONGSVTCFGEADQCAACAHYKDPFCVVAR 600
QY 600 CPGSVPRDLSTYPIWKFPDEEGICQPCPINCTHSCVDLDERGCPAQRASPV 652
DB 600 CPGSVPRDLSTYPIWKFPDEEGICQPCPINCTHSCVDLDERGCPAQRASPV 653
```

RESULT 14  
US-09-821-883-3  
Sequence 3, Application US/09821883  
Patent No. US20020061310A1  
GENERAL INFORMATION:  
APPLICANT: Laus, Reiner  
APPLICANT: Vidovic, Damir  
APPLICANT: Graddis, Thomas  
TITLE OF INVENTION: Compositions and Methods for Dendritic  
FILE REFERENCE: 7636-0022.30  
CURRENT APPLICATION NUMBER: US/09/821,883  
CURRENT FILING DATE: 2001-03-30  
PRIOR APPLICATION NUMBER: US 60/193,504  
PRIOR FILING DATE: 2000-03-30  
NUMBER OF SEQ ID NOS: 30  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 3  
LENGTH: 564  
TYPE: PRT



ORGANISM: Artificial Sequence  
 FEATURE:  
 OTHER INFORMATION: HER500\* construct  
 US-09-821-883-3

Query Match 50.9%; Score 2585; DB 10; Length 564;  
 Best Local Similarity 56.7%; Pred. No. 5,6e-142;  
 Matches 512; Conservative 3; Mismatches 6; Indels 382; Gaps 2;

16 LPPGASTVCTGTDKMLRLPASPEHLDMLRLYGGCOVVOGNLELTLPNTASISFQ 75  
 35 LARGAASVCTGTDKMLRLPASPEHLDMLRLYGGCOVVOGNLELTLPNTASISFQ 94  
 76 DIOEVGYVLIANOVROYPLQRLRVGTOLFEEDNYALAVLDNGPLNNTPTVGTASPG 135  
 95 DIOEVGYVLIANOVROYPLQRLRVGTOLFEEDNYALAVLDNGPLNNTPTVGTASPG 154  
 136 GLREIQLRSLEILKGVLIQRPOLCYODTILMKDIFHKNNQALTLIDTNSRACHPC 195  
 155 GLREIQLRSLEILKGVLIQRPOLCYODTILMKDIFHKNNQALTLIDTNSRACHPC 214  
 196 SPWCKSGRWGSSSEDCQSLTRVTCAGGACGKGRPLPTDCHEQCAAGCTGPKHSCLAC 255  
 215 SPWCKSGRWGSSSEDCQSLTRVTCAGGACGKGRPLPTDCHEQCAAGCTGPKHSCLAC 274  
 256 LHPNHSIGICELHCPALVTYNTDFFESMPNDEGRYTFGASCVTACPYNYLSTVGSCTIYC 315  
 275 LHPNHSIGICELHCPALVTYNTDFFESMPNDEGRYTFGASCVTACPYNYLSTVGSASIT- 333  
 316 PLHNEVTAEDGTORCEKSKPCARVCYGLGMEHLREVRAVTSANIOEPAGCKKIFGSLA 375  
 334 ----- 333  
 376 FLPESTGDPASNTAPLQPBQLOVFETLEITGYLYISAMPDSLPLSVFQNLQVIRGR 435  
 334 -----NEFKL----- 338  
 436 LHNQAVSLTLOGIGISWLGSLRLSRLSGSLALIHNTHLICFVHTVPWDLFRPHQALLH 495  
 339 ----- 338  
 496 TANRPEDECVGEGACIACIACIACIACIACIACIACIACIACIACIACIACIACIACIAC 555  
 339 ----- 338  
 556 NARHCLPCHPECPQNGSVTCFGRPADQCVACAHYKDRPFVYARCPGSKVYKPDLSYMPIMK 615  
 339 ----- 338  
 616 FDEBEGACQPCPICTHSCVLDLDDKCPADQASPLTSONEDLGPAASPLDSTFYRSLLED 675  
 339 ----- 338  
 676 DMGDLVDAEYLVPQGFPCDPAPGAGMHHRRSSSTRSGGDLTLGLPESEEAR 735  
 339 -----GAGMHHRRSSSTRSGGDLTLGLPESEEAR 372  
 736 RSPPLASBEGAGSDVFDGLGMAKGLQSLPTHDPSPLORYSEDPVPLPSETDGVAPL 795  
 373 RSPPLASBEGAGSDVFDGLGMAKGLQSLPTHDPSPLORYSEDPVPLPSETDGVAPL 432  
 796 TCSPOPEYVNOPDVPRQPPSPREGPLPARPAGATLERAKTISPCNGVYKDVFAFGAV 855  
 433 TCSPOPEYVNOPDVPRQPPSPREGPLPARPAGATLERAKTISPCNGVYKDVFAFGAV 492  
 856 ENPEYTLPOGGAPOPHRPAFPAFNDLYYWDOPPERGAPSTPKGTPTAENPEYTL 915  
 493 ENPEYTLPOGGAPOPHRPAFPAFNDLYYWDOPPERGAPSTPKGTPTAENPEYTL 552  
 916 DVP 918  
 553 DVP 555

## RESULT 15

US-09-821-883-4  
 Sequence 4, Application US/09821883

Patent No. US20020061310A1

GENERAL INFORMATION:

APPLICANT: Laus, Reiner

APPLICANT: Vidovic, Damir

APPLICANT: Graddis, Thomas

TITLE OF INVENTION: Compositions and Methods for Dendritic

TITLE OF INVENTION: Cell-Based Immunotherapy

FILE REFERENCE: 7636-0022.30

CURRENT APPLICATION NUMBER: US/09/821,883

PRIOR FILING DATE: 2001-03-30

PRIOR FILING DATE: 2000-03-30

NUMBER OF SEQ ID NOS: 30

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 4

LENGTH: 697

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: HER500\*-rsm-CSF construct

US-09-821-883-4

Query Match 50.9%; Score 2585; DB 10; Length 697;  
 Best Local Similarity 56.7%; Pred. No. 7,2e-142;  
 Matches 512; Conservative 3; Mismatches 6; Indels 382; Gaps 2;

16 LPPGASTVCTGTDKMLRLPASPEHLDMLRLYGGCOVVOGNLELTLPNTASISFQ 75  
 35 LARGAASVCTGTDKMLRLPASPEHLDMLRLYGGCOVVOGNLELTLPNTASISFQ 94  
 76 DIOEVGYVLIANOVROYPLQRLRVGTOLFEEDNYALAVLDNGPLNNTPTVGTASPG 135  
 95 DIOEVGYVLIANOVROYPLQRLRVGTOLFEEDNYALAVLDNGPLNNTPTVGTASPG 154  
 136 GLREIQLRSLEILKGVLIQRPOLCYODTILMKDIFHKNNQALTLIDTNSRACHPC 195  
 155 GLREIQLRSLEILKGVLIQRPOLCYODTILMKDIFHKNNQALTLIDTNSRACHPC 214  
 196 SPWCKSGRWGSSSEDCQSLTRVTCAGGACGKGRPLPTDCHEQCAAGCTGPKHSCLAC 255  
 215 SPWCKSGRWGSSSEDCQSLTRVTCAGGACGKGRPLPTDCHEQCAAGCTGPKHSCLAC 274  
 256 LHPNHSIGICELHCPALVTYNTDFFESMPNDEGRYTFGASCVTACPYNYLSTVGSCTIYC 315  
 275 LHPNHSIGICELHCPALVTYNTDFFESMPNDEGRYTFGASCVTACPYNYLSTVGSASIT- 333  
 316 PLHNEVTAEDGTORCEKSKPCARVCYGLGMEHLREVRAVTSANIOEPAGCKKIFGSLA 375  
 334 ----- 333  
 376 FLPESTGDPASNTAPLQPBQLOVFETLEITGYLYISAMPDSLPLSVFQNLQVIRGR 435  
 334 -----NEFKL----- 338  
 436 LHNQAVSLTLOGIGISWLGSLRLSRLSGSLALIHNTHLICFVHTVPWDLFRPHQALLH 495  
 339 ----- 338  
 496 TANRPEDECVGEGACIACIACIACIACIACIACIACIACIACIACIACIACIACIACIAC 555  
 339 ----- 338  
 556 NARHCLPCHPECPQNGSVTCFGRPADQCVACAHYKDRPFVYARCPGSKVYKPDLSYMPIMK 615  
 339 ----- 338  
 616 FDEBEGACQPCPICTHSCVLDLDDKCPADQASPLTSONEDLGPAASPLDSTFYRSLLED 675  
 339 ----- 338

Qy	676	DDMGDLVDAEYELVDPQGFCDPAPAGGCVVHRRHSSSTRBSGGDLTGLPEPSEAP	735
Dd	339	-----GAGGVVHRRHSSSTRBSGGDLTGLPEPSEAP	372
Qy	736	RSPLAPSEGAGSDVFDGDLGGAAGKGLQSLPTHPSPLORYSEDPTVPLPSETDGYAAPL	795
Dd	373	RSPLAPSEGAGSDVFDGDLGGAAGKGLQSLPTHPSPLORYSEDPTVPLPSETDGYAAPL	432
Qy	766	TCSPOPEVYNPDVROPSPREGLPLAARPAGATLERPTLSPGKNGVYKDVFAFGAV	855
Dd	433	TCSPOPEVYNPDVROPSPREGLPLAARPAGATLERATLTSBKNGVYKDVFAFGAV	492
Qy	856	ENPELTFPOGGAQDPHPPPASFAFDNLVYWDPPERCAPSTFGKPTAENPEYLG	915
Dd	493	ENPELTFPOGGAQDPHPPPASFAFDNLVYWDPPERCAPSTFGKPTAENPEYLG	552
Qy	916	DVP 918	
Dd	553	DVP 555	

Time : 22.7057 secs  
Job completed: January 13, 2003, 14:50:16

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 13, 2003, 14:44:55 : Search time 22.5383 Seconds  
(without alignments)  
3919.881 Million cell updates/sec

Title: US-09-854-356-6

Perfect score: 5078

Sequence: 1 MELALCRWGLLALLPPGA.....TFKGNPTAENPEYGLDVPV 919

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4900	96.5	1255	1 A24571	protein-tyrosine k
2	4134	81.4	1260	1 TVKTNV	protein-tyrosine k
3	4122.5	81.2	1254	2 I48161	p-185 precursor
4	1645	32.4	1210	1 GQHUE	epidermal growth f
5	1612	31.7	1210	2 A53183	epidermal growth f
6	1592.5	31.4	1223	1 TVCHLV	epidermal growth f
7	1536	30.2	644	2 A36325	epidermal growth f
8	1528.5	30.1	1308	2 A47253	epidermal growth f
9	1435	28.3	1342	2 A36223	kinase-related tra
10	1351.5	26.6	1339	2 JC4387	epidermal growth f
11	1313.5	25.9	1166	1 S06142	protein-tyrosine k
12	1155	22.7	527	2 A42032	epidermal growth f
13	1014.5	20.0	843	2 A27131	epidermal growth f
14	718.5	14.1	1323	2 E88257	protein-tyrosine k
15	718.5	14.1	1374	2 S70712	protein-tyrosine k
16	693	13.6	1330	1 GQFE	epidermal growth f
17	669.5	13.2	1369	2 S70713	protein-tyrosine k
18	621	12.2	1717	1 A45558	epidermal growth f
19	384	7.6	1363	2 T43220	insulin-like growt
20	376.5	7.4	1300	2 A36502	insulin receptor-t
21	370	7.3	1382	1 INHUR	insulin receptor p
22	363.5	7.2	366	2 D45558	epidermal growth f
23	362.5	7.1	1383	2 A36080	insulin receptor p
24	361	7.1	1383	2 A34157	insulin receptor p
25	353	7.0	1477	2 T18534	protein-tyrosine k
26	348.5	6.9	540	2 B47417	insulin receptor-t
27	338	6.7	1268	2 B36502	insulin receptor-t
28	331	6.5	333	2 B45558	epidermal growth f
29	331	6.5	342	2 C45558	epidermal growth f

30	326	6.4	1390	2 T30346	insulin receptor -
31	324	6.4	1607	2 T43212	insulin-like growt
32	316	6.2	1371	2 A33837	insulin-like growt
33	311.5	6.1	1367	1 TGHUR1	insulin-like growt
34	307	6.0	2101	2 S57245	insulin receptor (
35	307	6.0	2148	1 A56081	insulin receptor k
36	299.5	5.9	698	1 TVFVLV	protein-tyrosine k
37	267.5	5.3	1846	2 T42047	insulin receptor h
38	254.5	5.0	604	1 TVYUH	insulin receptor
39	244.5	4.8	1548	2 S34583	protein-tyrosine k
40	240	4.7	1299	2 T43251	serine proteinase
41	235	4.6	329	2 A48805	furin (EC 3.4.21.7
42	226.5	4.5	183	2 JH0803	insulin-like growt
43	208	4.1	1680	2 A43434	tyrosine kinase re
44	200	3.9	545	2 S00727	furin (EC 3.4.21.7
45	198.5	3.9	962	2 JC3571	kinase-related tra
					subtilisin-like pr

#### ALIGNMENTS

RESULT 1  
A24571  
protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human  
N:Alternate names: c-erb-B-2 protein precursor; kinase-related transforming protein e  
C:Species: Homo sapiens (man)  
C:Date: 25-Oct-1987 #sequence-revision 06-Dec-1996 #text-change 11-Jun-1999  
C:Accession: A24571; A25491; A44188; B44188; I59509; I57622  
R:Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; Saito, T  
Nature 319, 230-234, 1986  
A>Title: Similarity of protein encoded by the human c-erb-B-2 gene to epidermal growth  
A:Reference number: A24571; MUID:86118663; PMID:3003577  
A:Accession: A24571  
A:Molecule type: mRNA  
A:Residues: 1-1255 <YAM>  
A:Cross-References: GB:X03363; NID:931197; PIDN:CAA27060.1; PID:931198  
R:Semba, K.; Kamata, N.; Toyoshima, K.; Yamamoto, T.  
Proc. Natl. Acad. Sci. U.S.A. 82, 6497-6501, 1985  
A>Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/epid  
A:Reference number: A25491; MUID:86016729; PMID:2995967  
A:Accession: A25491  
A:Molecule type: DNA  
A:Residues: 737-1031 <SEM>  
A:Cross-References: GB:M11767; NID:9182163; PIDN:AAA35808.1; PID:9553282  
R:Cousens, L.; Yang-Feng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seeburg  
Science 230, 1132-1139, 1985  
A>Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares chro  
A:Reference number: A44188; MUID:86070181; PMID:2999974  
A:Accession: A44188  
A:Molecule type: DNA  
A:Residues: 740-910 <COU>  
A:Cross-References: GB:M12036; NID:9183988; PIDN:AAA35978.1; PID:9183989  
A:Accession: B44188  
A:Molecule type: mRNA  
A:Residues: 1-517, 'RALT', 522, 'S', 524-654, 'V', 656-1169, 'A', 1171-1255 <COU2>  
A:Cross-References: GB:M11730; NID:9183986  
R:King, C.R.; Kraus, M.H.; Aaronson, S.A.  
Science 229, 974-976, 1985  
A>Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma.  
A:Reference number: I59509; MUID:85272597; PMID:2992089  
A:Accession: I59509  
A:Molecule type: DNA  
A:Status: translated from GB/EMBL/DBJ  
A:Residues: 832-909 <REX>  
A:Cross-References: GB:L29395; NID:9459807; PIDN:AAA35809.1; PID:9459808  
R:Tal, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Givol, D.  
Mol. Cell. Biol. 7, 2597-2601, 1987  
A>Title: Human HER2 (neu) promoter: evidence for multiple mechanisms for transcriptio  
A:Reference number: I57622; MUID:87286898; PMID:3039351  
A:Accession: I57622  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-191 <TAU>

A:Cross-references: GB:M16792; NID:g183983; PIDN:AAA56637.1; PID:g553332  
 C:Comment: Amplification and overexpression of this erbB-related gene occurs in about 30  
 C:Genetics:  
 A:Gene: GDB:ERB2; NGL; NEU; HER-2  
 A:Cross-references: GDB:120613; OMIM:164870  
 A:Map position: 17q21.1-17q21.1  
 A:Introns: 25/1; 75/3; 147/1; 883/3  
 A>Note: the list of introns is incomplete  
 C:Function:  
 A:Description: catalyzes the phosphorylation of a peptidyl tyrosine residue by ATP  
 C:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho  
 kinase  
 F:1-21/Domain: signal sequence #status predicted <SIG>  
 F:22-155/Product: protein-tyrosine kinase erbB2 #status predicted <MAT>  
 F:22-653/Domain: extracellular #status predicted <EXT>  
 F:70-304/Domain: EGF receptor extracellular domain repeat <EB1>  
 F:395-605/Domain: EGF receptor extracellular domain repeat <EB2>  
 F:654-675/Domain: transmembrane #status predicted <TM>  
 F:676-1255/Domain: intracellular #status predicted <INT>  
 F:718-983/Domain: protein kinase homology <KIN>  
 F:726-734/Region: protein kinase ATP-binding motif  
 F:66,124,187,259,530,571,639/Binding site: carbohydrate (Asn) (covalent) #status predict  
 F:686/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted  
 F:759/Active site: lys #status predicted  
 F:1139,1221,1222,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

Query Match 96.5%; Score 4900; DB 1; Length 1255;  
 Best Local Similarity 73.2%; Pred. No. 3.3e-263;  
 Matches 919; Conservative 0; Mismatches 0; Indels 336; Gaps 1;

OY 1 METALCRWGLLALLPRGAASVCTGTDKRLRSPETHDMLRLHYOGCVVQGNL 60  
 DB 1 MELALCRWGLLALLPRGAASVCTGTDKRLRSPETHDMLRLHYOGCVVQGNL 60  
 OY 61 ELYTLPTNASISFQDIOEVGVYLIANQVROYPLQRLTVRGSTOLFEDNYALAVDNG 120  
 DB 61 ELYTLPTNASISFQDIOEVGVYLIANQVROYPLQRLTVRGSTOLFEDNYALAVDNG 120  
 OY 121 DPLNNTPTVTCASPGRLRELOLRSLTEILKGVLIQRPOLCYDTILMKDIFKNNOLA 180  
 DB 121 DPLNNTPTVTCASPGRLRELOLRSLTEILKGVLIQRPOLCYDTILMKDIFKNNOLA 180  
 OY 181 LTLIDTNSRACHPCSPCKSGRCWSESSDQSLFTVTCAGGACRCKGPLPTDCHEQC 240  
 DB 181 LTLIDTNSRACHPCSPCKSGRCWSESSDQSLFTVTCAGGACRCKGPLPTDCHEQC 240  
 OY 241 AAGCTGPRHSDCLACLNHNSGICELHCPALVTYNTDFFESMPNEGRYTGASCVTACP 300  
 DB 241 AAGCTGPRHSDCLACLNHNSGICELHCPALVTYNTDFFESMPNEGRYTGASCVTACP 300  
 OY 301 YNYLSTVGSCSTIVCPRLNQETAEADGTQRCCKSPCARCYGIGMHLREVRVAVTSAN 360  
 DB 301 YNYLSTVGSCSTIVCPRLNQETAEADGTQRCCKSPCARCYGIGMHLREVRVAVTSAN 360  
 OY 361 IQEFAAGCKKIFGSLAFPEFSGDPASNTAPLQEQLOVFETLEETGYLIASMPDSL 420  
 DB 361 IQEFAAGCKKIFGSLAFPEFSGDPASNTAPLQEQLOVFETLEETGYLIASMPDSL 420  
 OY 421 DLSVFNQVYRGRILNNGASITLQIGISWLGRLSRISRELGSGALLHNHTHLCFVTV 480  
 DB 421 DLSVFNQVYRGRILNNGASITLQIGISWLGRLSRISRELGSGALLHNHTHLCFVTV 480  
 OY 481 PMDQLFRPHOALLHTANRPEDECVGEGILACHOLCARGCHMGPGPTQCVNCSQFLRGQEC 540  
 DB 481 PMDQLFRPHOALLHTANRPEDECVGEGILACHOLCARGCHMGPGPTQCVNCSQFLRGQEC 540  
 OY 541 VEECRVLQGLRREYVNAHCLPCHECOPONGSVTCFCEPADQCVACAHYKDRPFCVARC 600  
 DB 541 VEECRVLQGLRREYVNAHCLPCHECOPONGSVTCFCEPADQCVACAHYKDRPFCVARC 600  
 OY 601 PSGVKPDLSTYPIWKFPDEBACOPCPINCTHSCVDDLDDKCPAEQASPLTSLTISAVYG 660  
 DB 601 PSGVKPDLSTYPIWKFPDEBACOPCPINCTHSCVDDLDDKCPAEQASPLTSLTISAVYG 660

DB 601 PSGVKPDLSTYPIWKFPDEBACOPCPINCTHSCVDDLDDKCPAEQASPLTSLTISAVYG 660  
 OY 654 ----- 653  
 DB 661 ILVVLVGVVGLIKRRQOKIRKYTMRLLOETELVEPLTPSGAMNQAMRIKETEEL 720  
 OY 654 ----- 653  
 DB 721 RKKVKLGSAGFTYKGIWIPDGENVKIPVAIKYLRNTPSKANKETLDEAYVAGVGP 780  
 OY 654 ----- 653  
 DB 781 YVSRLLGICLSTYQVLTQMLPFGCCLLDHYRENKRGSDLLMWCQIAKMSYLEDVR 840  
 OY 654 ----- 653  
 DB 841 LVHRDLAARNLVKSPNHVKITDFGLARLIDIDETEHADGKVPYIKMALESILRRFT 900  
 OY 654 ----- 653  
 DB 901 HQSDWSYGVYTWELMTFGAKPYDGIPIAREIPIDLLEKGERLPQPICTIDVYIMVKCM 960  
 OY 654 ----- 684  
 DB 961 IDSCRRFRFLVSEFSRMARDPQRFVYIQNEDLGPASPLDSTYRSLLEDMDKGLVDA 1020  
 OY 685 EELVLPQGFPCDPAPGAGGVHRRHRSSTRSGGDLTLGLPSEEARPSPLASEG 744  
 DB 1021 EELVLPQGFPCDPAPGAGGVHRRHRSSTRSGGDLTLGLPSEEARPSPLASEG 1080  
 OY 745 AGSVFPGDGLMGAAKGLQSLPTHDPSPLQYSSDPVPLPSEMDGVYAPLTCSPQPEYV 804  
 DB 1081 AGSVFPGDGLMGAAKGLQSLPTHDPSPLQYSSDPVPLPSEMDGVYAPLTCSPQPEYV 1140  
 OY 805 NOPDVROPSPRSGPLPAARPAATLERPTLSPKNGVYKDVFAFGAENPEYLTPO 864  
 DB 1141 NOPDVROPSPRSGPLPAARPAATLERPTLSPKNGVYKDVFAFGAENPEYLTPO 1200  
 OY 865 GGAAPQHPAPPAFPAFDNLYWDDPPERGAPSTFKGPTAENPEYLGIDVY 919  
 DB 1201 GGAAPQHPAPPAFPAFDNLYWDDPPERGAPSTFKGPTAENPEYLGIDVY 1255

RESULT 2  
 TVRKND  
 protein-tyrosine kinase (EC 2.7.1.112) neu precursor - rat  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 31-Dec-1988 #sequence\_revision 31-Dec-1988 #text\_change 11-Jun-1999  
 C:Accession: A24562; A61204  
 R:Barigmann, C.I.; Hung, M.C.; Weinberg, R.A.  
 Nature 319, 226-230, 1986  
 A:Title: The neu oncogene encodes an epidermal growth factor receptor-related protein  
 A:Reference number: A24562; MUID:86118662; PMID:3945311  
 A:Accession: A24562  
 A:Molecule type: mRNA  
 A:Residues: 1-1260 <BAR>  
 A:Cross-references: EMBL:X03362; NID:956745; PIDN:CAAZ2059.1; PID:956746  
 R:Masui, T.; Mann, A.M.; Macatee, T.L.; Garland, E.M.; Okamura, T.; Smith, R.A.; Cohe  
 Carcinogenesis 12, 1975-1978, 1991  
 A:Title: Direct DNA sequencing of the rat neu oncogene transmembrane domain reveals n  
 2-thiazolyl]formamide or N-methyl-N-nitrosourea  
 A:Reference number: A61204; MUID:92035293; PMID:1682063  
 A:Accession: A61204  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 637-663, 'V', 665-702 <MAS>  
 A>Note: authors translated the codon GCA for residue 25 as Val  
 C:Genetics:  
 A:Gene: neu  
 C:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phos  
 F:1-19/Domain: signal sequence #status predicted <SIG>  
 F:120-1260/Product: protein-tyrosine kinase neu #status predicted <MAT>

F:658-680/Domain: transmembrane #status predicted <TMN>

F:723-988/Domain: protein kinase homology <KIN>

F:731-739/Region: protein kinase ATP-binding motif

F:711,191,263,535,576,634/Binding site: carbohydrate (Asn) (covalent) #status predicted

F:691/Binding site: phosphate (Thr) (covalent) #status predicted

F:758/Active site: Lys #status predicted

F:882,1227,1253/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 81.4%; Score 4134; DB 1; Length 1260;

Best Local Similarity 62.1%; Pred. No. 7.1e-221;

Matches 761; Conservative 42; Mismatches 96; Indels 338; Gaps 3;

```

OY 1 METLAALCRMGILLALLPRGASSTOVCTGTMKRLPRASPETHIDMLRHLYOGCCVVOGNTL 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4 MELAAACRMGFLALLPRPGIAGTGYCTGTMKRLPRASPETHIDMLRHLYOGCCVVOGNTL 63
OY 61 ELYTLPNASTLFLQDIOEVGYVLLAHNOVROVPLORLRIYVGTQLFEDNYALAVLDNG 120
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 64 ELTYVPANASTLFLQDIOEVGYVLLAHNOVKRVPLORLRIYVGTQLFEDKYALAVLDNR 123
OY 121 DPLNNTPTV-GASPGGLRELQLSLTEILKGYLIORNPOLCYQDTILMKDIFHKNNQL 179
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 124 DPODNVAASPTGRTPEGLRELQLSLTEILKGYLIORNPOLCYQDMVYLKWDVFRKNNQL 183
OY 180 ALTLIDNRSRACHPCSPMCKSGSRMGSESDCOSLRTVYACAGCARGKPLPTDCHEQ 239
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 184 APVDIDNRSRACHPCSPMCKSGSRMGSESDCOSLRTVYACAGCARGKPLPTDCHEQ 243
OY 240 CAAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDPEFSMPNBRGYTFGASCVTAC 299
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 244 CAAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDPEFSMPNBRGYTFGASCVTTC 303
OY 300 PYNYLSTDVGSCTLVCPILNNOEVTAEEDGTORCEKSCPKARVCYGLGMEHLREVRAYTSA 359
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 304 PYNYLSTDVGSCTLVCPILNNOEVTAEEDGTORCEKSCPKARVCYGLGMEHLREVRAYTSD 363
OY 360 NIOEFAGCKKIFGSLAFLRPSFGDDPSANTAPLOPEQLQVETLEITGGLYISAMPDSL 419
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 364 NIOEFAGCKKIFGSLAFLRPSFGDDPSANTAPLOPEQLQVETLEITGGLYISAMPDSL 423
OY 420 PDLVSFONLQVINGRILHNGAYSLLTQGLGISWLGKLSRELSSGALLHHNTHLCFVHT 479
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 424 RDLVSFONLQVINGRILHNGAYSLLTQGLGISWLGKLSRELSSGALLHHNTHLCFVHT 483
OY 480 VPMDOQLFRNHQALHNGAYSLLTQGLGISWLGKLSRELSSGALLHHNTHLCFVHT 538
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 484 VPMDOQLFRNHQALHNGAYSLLTQGLGISWLGKLSRELSSGALLHHNTHLCFVHT 543
OY 539 ECVBECHVLOGLPREVYNAHRLCPARPECOPONGSVTCFGEADQCYAACHYKDPPECVA 598
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 544 ECVBECHVLOGLPREVYNAHRLCPARPECOPONGSVTCFGEADQCYAACHYKDPPECVA 603
OY 599 RCBGKVPDLSTYPIWKFPEDEGACQPCPINCSTHSCVDLDDKGCAPABORASPLT----- 652
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 604 RCBGKVPDLSTYPIWKFPEDEGACQPCPINCSTHSCVDLDDKGCAPABORASPLT----- 663
OY 653 ----- 652
DB 664 EGYVLLTILVNVVGLIKRRRQKIRKTYMRLLQETELVEPLTPSGAMPNOAQMRLIKET 723
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 653 ----- 652
DB 724 ELKRVKYLSSGAGFYVKGIMIPDGENVKIPVAIKVIRENTSPRANKELIDEAYVMAVG 783
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 653 ----- 652
DB 784 SPYVSRLLIGLITSTVQLVTLMPYGLLDHVEHNRGLSGODLLMNCVOIAKGMSTLED 843
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 653 ----- 652
DB 844 VRLVHRDLAARNVLVSPNHVKTITDEGLARLDIDETEHADGKVPKIMMALESILRR 903
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 653 ----- 652

```

```

DB 904 FTHOSDVWSGVYVWEIMLTGAKRKYDGIAPAREIPDLLEKGERLPQPPICITIDVYMIWVC 963
OY 653 ----- 682
DB 964 WMIDSECRPRFRELSEFSRWARDPQRFVVIYQNEIDLSPSPMDSTFYRSLLEDDMDGLV 1023
OY 683 DAEEYLVPOGGFRCPPDPAPAGAGVHHRRSSSTRSGGDLTIGLESEDEAPRSLAPS 742
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1024 DAEEYLVPOGGFRCPPDPAPAGAGVHHRRSSSTRSGGDLTIGLESEDEAPRSLAPS 1083
OY 743 EGASDVFDGDLGKAAGKQSLPTHDPSLQRYSEPTVPLPSETDGYVAAPLTCQPOE 802
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1084 EGASDVFDGDLGKAAGKQSLPTHDPSLQRYSEPTVPLPSETDGYVAAPLTCQPOE 1143
OY 803 YVNOQVYRQPPSPREGPLPAPRAGATLERPKTSLPGKNGVYKDVAFGAVENPEYLT 862
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1144 YVNOQVYRQPPSPREGPLPAPRAGATLERPKTSLPGKNGVYKDVAFGAVENPEYLT 1203
OY 863 POGGAAPQHPHPPAFSPAFNDLYWQDDPREKAPSTFEGTPTAENPEYLTGLDVPY 919
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1204 PREGTAAPHPSPAFSPAFNDLYWQDDPREKAPSTFEGTPTAENPEYLTGLDVPY 1260

```

RESULT 3

148161

P-185 precursor - golden hamster

C:Species: Mesocricetus auratus (golden hamster)

C:Date: 02-Jul-1996 #sequence\_rev1sion 02-Jul-1996 #text\_change 18-Jun-1999

C:Accession: I48161

R:Nakamura, T.; Ushijima, T.; Ishizaka, Y.; Nagao, M.; Arai, M.; Yamazaki, Y.; Ishika

Gene 140, 251-255, 1994

A:Title: Cloning and activation of the Syrian hamster neu proto-oncogene.

A:Reference number: I48161; M0ID:94193007; PMID:7908275

A:Accession: I48161

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1254 <RES>

A:Cross-references: GB:D16295; NID:9493236; PIDD:BA03801.1; PID:9747595

C:Genetics:

A:Gene: neu

C:Superfamily: epidermal growth factor receptor; protein kinase homology

C:Keywords: ATP

F:718-983/Domain: protein kinase homology <KIN>

F:726-734/Region: protein kinase ATP-binding motif

Query Match 81.2%; Score 4122.5; DB 2; Length 1254;

Best Local Similarity 61.8%; Pred. No. 3.1e-220;

Matches 776; Conservative 49; Mismatches 93; Indels 337; Gaps 2;

```

OY 1 METLAALCRMGILLALLPRGASSTOVCTGTMKRLPRASPETHIDMLRHLYOGCCVVOGNTL 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1 MELAAACRMGILLALLSPGASSTOVCTGTMKRLPRASPETHIDMLRHLYOGCCVVOGNTL 60
OY 61 ELYTLPNASTLFLQDIOEVGYVLLAHNOVROVPLORLRIYVGTQLFEDNYALAVLDNG 120
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 61 ELYTLPNASTLFLQDIOEVGYVLLAHNOVKRVPLORLRIYVGTQLFEDKYALAVLDNR 120
OY 121 DPLNNTPTV-GASPGGLRELQLSLTEILKGYLIORNPOLCYQDTILMKDIFHKNNQL 180
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 121 DPLNNTPTV-GASPGGLRELQLSLTEILKGYLIORNPOLCYQDTILMKDIFHKNNQL 180
OY 181 LTLIDNRSRACHPCSPMCKSGSRMGSESDCOSLRTVYACAGCARGKPLPTDCHEQ 240
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 181 LTLIDNRSRACHPCSPMCKSGSRMGSESDCOSLRTVYACAGCARGKPLPTDCHEQ 240
OY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDPEFSMPNBRGYTFGASCVTAC 300
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDPEFSMPNBRGYTFGASCVTTC 300
OY 301 YNYLSTDVGSCTLVCPILNNOEVTAEEDGTORCEKSCPKARVCYGLGMEHLREVRAYTSAN 360
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 301 YNYLSTDVGSCTLVCPILNNOEVTAEEDGTORCEKSCPKARVCYGLGMEHLREVRAYTSAN 360

```

QY	361	IOEFNGCKKIFGSLAFLESEFGDEASNTAPLOPQLOVFELEETITGVLITSAMPDLP	420
Db	361	IOEFNGCKKIFGSLAFLESEFGDNSSGIAPLTPBQLOVFELEETITGVLITSAMPDLP	420
QY	421	DLVSFQNTQVIRGRILTHNGAVSLTLOGIGISWLGISRLSRELGSGIALIHNHTHLCEVTV	480
Db	421	DLVSFQNTQVIRGRILTHNGAVSLTLOGIGISWLGISRLSRELGSGIALIHNHTHLCEVTV	480
QY	481	FMDOLEFRPHOALLHTARPEDECEYGBELACHOICARHCNMGPTQOCVNSQPLRGEC	540
Db	481	FMDOLEFRPHOALLHTARPEDECEYGBELACHOICARHCNMGPTQOCVNSQPLRGEC	540
QY	541	VEECFVLOGLEPREYVNAHCLPCHEPCOPONGSVTCFPEPDQCAACHYKODPECVARC	600
Db	541	VKECVWKGLEPREYVNAHCLPCHEPCOPONGSVTCFPEPDQCAACHYKODPECVARC	600
QY	601	PSGVPEDLSTYMPYKFPDEBGACOPCPICTHSCVDLDDKCCPAEORASPLTS-----	653
Db	601	PSGVPEDLSTYMPYKFPDEBGACOPCPICTHSCVDLDDKCCPAEORASPLTSIVTVYG	660
QY	654	-----	653
Db	661	ILFLVIGVYVVGILIKRRRQKIRKTYMRLLQETELVEPLTPSGAMPQAOAMRIKETEL	720
QY	654	-----	653
Db	721	RKVKVLGSGAGTVYKGIWIDGENVKIPVAIKVLRNTSPKANKEILDEAYVYVWAGLSP	780
QY	654	-----	653
Db	781	YVSRLLGICLSTYQVLYQMLPYGCLLDHVBHNRGLCSQDLLMVCVOIAKMGSLIEDVR	840
QY	654	-----	653
Db	841	LVHRDLARNVLYKSPNHVKITDFGLARLDIDETRYHADGKVPYKIALESILRRFT	900
QY	654	-----	653
Db	901	HOSDVMSYGVTVWELMTFGARPYDGIIPAREIPDLLEKGERLPQPICTIDVYIMVYKCM	960
QY	654	-----QNEIDLGPASPLDSTFYRSLLEDDMDGLDYDA	684
Db	961	IDSECRPFRELYSEFSRMDARDPQRFVYIQNEIDLGPASPLDSTFYRSLLEDDMDGLDYDA	1020
QY	665	EEXLYPQOGFCPPDAPAGAGVHHRRSSSTRSGGDLVTLGLEBSEADPRSLABSEG	744
Db	1021	EEXLYPQOGFCPPDAPAGAGVHHRRSSSTRSGGDLVTLGLEBSEADPRSLABSEG	1080
QY	745	AGSDVFDLDMGAKGQSLPTHDPSPLQRYSEDPVPLPSMDNGVYVAPLTCSPQPEYV	804
Db	1081	AGSDVFEELMGAKGQKQOSISPRQLSPLOKRYSEDPVPLPTETDGYVAPLACSPQPEYV	1140
QY	805	NOPDVROPSPREGPLPAARPAAGATLERPXTLSPGKNGVYKDYAFAGAVENPEYLP	864
Db	1141	NQPEVRPOPPLTEBGPPLPVPBPAGATLERPXTLSPGKNGVYKDYVTFEGAVENPEYLPVR	1200
QY	865	GGAAPQHPHPAPAFSPAFDNLVYMDODPREBGAPOSTFGTGTAEKPEYLGIDVY	919
Db	1201	GGSSAQPH-PPALCPAFDNLVYMDODPREBGAPOSTFGTGTAEKPEYLGIDVY	1254

**RESULT 4**  
**QCODE**  
 epidermal growth factor receptor precursor - human  
 N:Contains: protein-tyrosine kinase (EC 2.7.1.112) erdb  
 C:Species: Homo sapiens (man)  
 C:Date: 15-Nov-1984 #sequence\_revision 27-Nov-1985 #text\_change 11-Jun-1999  
 C:Accession: A00641; A25772; S30024; A38672; A00644; A43615; A23062; A05281; A60143; A23  
 R:Rullrich, A.; Coussens, L.; Hayflick, J.S.; Dull, T.J.; Gray, A.; Tam, A.W.; Lee, J.; Y  
 rg, P.H.  
 Nature 309, 418-425, 1984  
 A:Title: Human epidermal growth factor receptor cDNA sequence and aberrant expression of

A:Reference number: A00641; MUID:84219729; PMID:6328312  
A:Accession: A00641  
A:Molecule type: mRNA  
A:Residues: 1-1210 <DL>  
A:Cross-references: EMBL:X00588; NID:g931113; PIDN:CAA5240.1; PID:g757924  
A:Note: the authors translated the codon AAG for residue 540 as Asn  
R:Smith, S.; Xu, Y.; Stratton, R.H.; Roe, B.A.; Mellino, G.T.; Pastan, I.  
Proc. Natl. Acad. Sci. U.S.A. 82, 4920-4924, 1985  
A:Title: Characterization and sequence of the promoter region of the human epidermal  
A:Reference number: A25772; MUID:85270438; PMID:2991899  
A:Accession: A25772  
A:Status: translation not shown  
A:Molecule type: DNA  
A:Residues: 1-29 <ISH>  
A:Cross-references: GB:M1234; NID:g181981; PIDN:AAA53270.1; PID:g553272  
R:Halley, J.; Whittle, N.; Bennett, P.; Kitchington, D.; Ullrich, A.; Waterfield, M.  
Oncogene Res. 1, 375-396, 1987  
A:Title: The human EGF receptor gene: structure of the 110 kb locus and identificatio  
A:Reference number: S30024; MUID:88217333; PMID:3329716  
A:Accession: S30024  
A:Molecule type: DNA  
A:Residues: 1-29 <HA>  
A:Cross-references: EMBL:X06370; NID:g31118; PIDN:CAA29668.1; PID:g31119  
R:Halley, J.D.; Waterfield, M.D.  
J. Biol. Chem. 266, 1746-1753, 1991  
A:Title: Contributory effects of de Novo transcription and premature transcript termi  
A:Reference number: A38672; MUID:91107677; PMID:1988448  
A:Accession: A38672  
A:Molecule type: DNA  
A:Residues: 1-29 <HAL>  
A:Cross-references: GB:X38425; NID:g181977; PIDN:AAA63171.1; PID:g553271  
A:Experimental source: carcinoma cell line A431-7  
R:Xu, Y.; Ishii, S.; Clark, A.C.L.; Sullivan, M.; Wilson, R.K.; Ma, D.P.; Roe, B.A.;  
Nature 309, 806-810, 1984  
A:Title: Human epidermal growth factor receptor cDNA is homologous to a variety of RN  
A:Reference number: A00642; MUID:84245835; PMID:6330563  
A:Accession: A00642  
A:Molecule type: mRNA  
A:Residues: RCAMRA, 150-187, 'KSYQAV', 195, 'W', 197, 'A', 199-222, 'S', 224-304, 'RA', 307-  
, 798-799, 'TD', 802-811, 'R', 813-942 <XUY>  
A:Experimental source: A431 human carcinoma cells, which have large numbers of EGF re  
R:Lin, C.R.; Chen, W.S.; Krutiger, W.; Stolarsky, L.S.; Weber, W.; Evans, R.M.; Verna,  
Science 224, 843-848, 1984  
A:Title: Expression cloning of human EGF receptor complementary DNA: gene amplificati  
A:Reference number: A43615; MUID:84196372; PMID:6326261  
A:Accession: A43615  
A:Molecule type: mRNA  
A:Residues: 713-964 <LIN>  
A:Experimental source: epidermoid carcinoma cell line A431  
R:Stimmen, F.A.; Gope, M.L.; Schultz, T.Z.; Wright, D.A.; Carpenter, G.; O'Malley, B.W.  
Biochem. Biophys. Res. Commun. 124, 125-132, 1984  
A:Reference number: A23062; MUID:85046483; PMID:6093780  
A:Accession: A23062  
A:Molecule type: mRNA  
A:Residues: 1028-1210 <SIM>  
R:Weber, W.; Gill, G.N.; Speiss, J.  
Science 224, 294-297, 1984  
A:Reference number: A05281; MUID:84172183; PMID:6324343  
A:Accession: A05281  
A:Molecule type: protein  
A:Residues: 25-30,'S',32-51,'454-467 <WEB>  
R:Russo, M.W.; Lukas, T.J.; Cohen, S.; Staros, J.V.  
J. Biol. Chem. 260, 5205-5208, 1985  
A:Title: Identification of residues in the nucleotide binding site of the epidermal g  
A:Reference number: A60143; MUID:85182650; PMID:2985580  
A:Accession: A60143  
A:Molecule type: protein  
A:Residues: 740-744, 'X', 746-747 <RUS>  
R:Mróczkowski, B.; Mosig, G.; Cohen, S.  
Nature 309, 270-273, 1984  
A:Title: ATP-stimulated interaction between epidermal growth factor receptor and supe  
A:Reference number: A38023; MUID:84191554; PMID:6325948  
A:Contents: annotation; receptor activity

A: Note: the EGF receptor (and other tyrosine kinases) can nick double-stranded DNA  
 R: Chen, W.S.; Lazar, C.S.; Lund, K.A.; Welsh, J.B.; Chang, C.P.; Walton, G.M.; Der, C.J.  
 Cell 59, 33-43, 1989  
 A: Title: Functional independence of the epidermal growth factor receptor from a domain I  
 A: Reference number: A33331; MIM:90003233; PMID:2709060  
 A: Contents: annotation; internalization signal  
 C: Comment: Binding of EGF to the receptor leads to internalization of the EGF-receptor C  
 C: Genes:  
 A: Gene: GDB:EGFR  
 A: Cross-references: GDB:120610; OMIM:131550  
 A: Map position: 7p12.3-7p12.1  
 C: Superfamily: epidermal growth factor receptor; protein kinase homology  
 C: Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phosph  
 F:1-24/Domain: signal sequence #status predicted <Sig>  
 F:25-1210/Product: EGF receptor #status predicted <MAT>  
 F:25-645/Domain: extracellular #status predicted <EXT>  
 F:75-300/Domain: EGF receptor extracellular domain repeat <EE1>  
 F:390-600/Domain: EGF receptor extracellular domain repeat <EE2>  
 F:646-668/Domain: transmembrane #status predicted <TM>  
 F:669-1210/Domain: intracellular #status predicted <INT>  
 F:710-925/Domain: protein kinase homology <KIN>  
 F:718-726/Region: protein kinase ATP-binding motif  
 F:999-1046/Region: coated-pit mediated internalization signal  
 F:1047-1210/Region: inhibitory  
 F:128-175,352,413,444,528,603/Binding site: carbohydrate (Asn) (covalent) #status predic  
 F:745/Active site: Lys #status experimental

Query Match 32.4%; Score 1645; DB 1; Length 1210;  
 Best Local Similarity 29.6%; Pred. No. 2,4e-83;  
 Matches 374; Conservative 140; Mismatches 309; Indels 442; Gaps 22;

11 LLLALLPAGAA--STGVCTGDMKLRPASEPETHLMDLRHLYGGCOVVOGNTLETYPFN 68  
 14 LLLALLPAGAA--STGVCTGDMKLRPASEPETHLMDLRHLYGGCOVVOGNTLETYPFN 73  
 69 ASLFLDIDQVQVYVLIANQVROYPLQRLIVRGTOLEEDNYALAVLDNGPLNNTTP 128  
 74 YDLSTFKTIOEVAGYVLIANTVERIPLENLQIRGMYYENSLAVALENTD----- 126  
 129 VFGASPGELRELQRLSTTEILKGVLQIRNPOLCYODTILMKDIFKNNOLATLTDNR 188  
 127 ---ANKTGLKELPMRNQLTEILHGAVRSPNPALCNVESIQMRDIVSDFLSNMSMDPN 183  
 189 SRACHPCSPKCSRCMGESSEDOQSITRTVCAGCA--RCKGPLPTCCHEGCAAGTGP 247  
 184 LGSQCQKCDPSPGSCGAGEENCQKRLTKLICAOQCSGRCGRKSPSCCHNCAAGCTGP 243  
 248 KHSDDLACTLHFNHSGICELCPALVTYNTDTFESMPREGRTFGASCVTACPYNYLSTD 307  
 244 RESDCLVCRKFRDEATCKDCPLMLXNPTTYQMDVNPBEKYSFGATCVAKCRPNYVTD 303  
 308 VGSCTLVCPILHNOEYTAEDGTQRCCKSCPCARVCYGLGMEHLREVAVTSAIQEAGC 367  
 304 HGSCVRCAGADSYEM--EEDVRRKCKCEGRCRCVNCINGIGEFKDSLSINATNKHKKN 362  
 368 KRIFGSLATPESFDGDRASNTAPLQEQVETLEITGLYLYISMPSLDLSFYQN 427  
 363 TTSISGDHLPLVAFRGDSFHTPLDQELDKTKVKEITGFLLIQAMPENRDLAFEN 422  
 428 LQYTRGILHNGAVSLTLQGLISWGLRLSRLRELSGLALHNNTHLCFVHTVPMQDLFR 487  
 423 LEIRGTRKHQGFSLAVVSLNTSLGRSLKEISDGDVILISGNKNLCYANTINMKKLF 482  
 488 NPHQALLHTANRDEDECVGGLACHOLCARGHCWGPPTQCVNCSQFLRQCEVEECRYL 547  
 483 TSGQKFTIISRGNSCKATGVCHALCSPEGCMGPEPRPCVSCRNVSRRGECVCKKL 542  
 548 QGLPREVYVNAHRLPCIRPEQOPONGSTVCGPRADQCVAAHAKDPFCVAPRSGKPD 607  
 543 EGPREFEVNSCIIQCHPELPLQAMNNTCTGRRPDNCIOCAHYIDGCHVKCTPAGVME 602  
 608 LSYMPIKPFDEGACOPCINCITNSCVDDDDGCPA----- 644

Db 603 NNTL-VMKYADAGVCHLCHPNCYTGCTGPGLEGCPTNGKIPSIATGVALLLLVVA 661  
 Qy 645 -----EQRASPULTSONE----- 656  
 Db 662 LGIGLFMRHRHIVKRTLRRLLOERLVEPLTFSGEAPNOALLRLKETEFKKIKVLGSG 721  
 Qy 657 ----- 656  
 Db 722 AFGTVYKGLMIPGEKKIVAIKELREATSPANKETLDEAVYMASVDNPHVCRLLGIC 761  
 Qy 657 ----- 656  
 Db 782 LFTSVQLITQLMFPGCLLDVYREHKVIGSQYLNNMCVQIAGKNYLEDRLVHDLAAR 841  
 Qy 657 -----DLG----- 659  
 Db 842 NVLVKTPQHVKTIDFGIAGLKGAEKEKYNAEGKVPIKMMALSIHRIYHOSDWSYG 901  
 Qy 660 -----PA----- 661  
 Db 902 VYWEIMTFGSKPYDGIPIASEISSLKGERLPQPICTIDVYIMVKKMIDADSHPKF 961  
 Qy 662 -----SPLDSTFYRLIEDDMDGDLVDAEYLVPOQ 692  
 Db 962 RELIETSKMARDPQRYLVIGDERMHLPSPTDSNFYRALMDEEDMDVDYDADEYLLPQ 1021  
 Qy 693 GFPCPPDPAAGGVNHHRRSSSTRSGGDLTLGLEPSEEARSPVLAISGASDVFDG 752  
 Db 1022 GFF-----SSPSTSRPTLLSLSTASN--NS 1045  
 Qy 753 DLGGAAGLQSLPTHPSPIQRYSEDPTVPLPSET--DGYVAPLQCSPOPEVVDNPDVR 810  
 Db 1046 TVACIDRNGLSQCPITKEDSLQRYSSDPICALIEDSLDITFL-----PYPETINO---- 1095  
 Qy 811 POPSPREGPLPAPAPAGATL-----ERPXTLSPGKGVYKDYFAFGAVENPEYL-TPQ 864  
 Db 1096 -----SVP-KRPAQSVQNPVYHMQPLNPAASRDPHYQD--PSTAVGNREYLMVTQ 1143  
 Qy 865 GGAAPDHPPPAPFADNLITYDQ-----DP-----PENGAPSTKGTPTAE 908  
 Db 1144 -----PTCVNSFSDSPAHMAQKSHQISLDNPYQODFPPEKAPKPIGIFKS-TAE 1193  
 Qy 909 NPEYL 913  
 Db 1194 NAEYL 1198

RESULT 5  
 A53183  
 epidermal growth factor receptor precursor - mouse  
 C:Species: Mus musculus (house mouse)  
 C:Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 18-Jun-1999  
 C:Accession: A53183; A43818; S24942; A28941; S45325; I49643  
 R:Luetteke, N.C.; Phillips, H.K.; Qiu, T.H.; Copeland, N.G.; Earp, H.S.; Jenkins, N.A  
 Genes Dev. 8, 399-413, 1994  
 A: Title: The mouse waved-2 phenotype results from a point mutation in the EGF recepto  
 A: Reference number: A53183; MIM:94170986; PMID:812525  
 A: Accession: A53183  
 A: Molecule type: mRNA  
 A: Residues: 1-1210 <LDB>  
 A: Cross-references: GB:U03425  
 R:Avioli, A.; Lax, I.; Ullrich, A.; Schlessinger, J.; Givol, D.; Morse, B.  
 Oncogene 6, 673-676, 1991  
 A: Title: Comparison of EGF receptor sequences as a guide to study the ligand binding  
 A: Reference number: A43818; MIM:91232866; PMID:2030916  
 A: Accession: A43818  
 A: Molecule type: mRNA  
 A: Residues: 1-714 <AVI>  
 A: Cross-references: GB:X59698  
 R:Elisinger, D.P.; Serretero, G.  
 submitted to the EMBL Data Library, June 1992  
 A: Reference number: S24942  
 A: Accession: S24942





A:Accession: A27720  
A:Molecule type: mRNA  
A:Residues: 1-1223 <LAX>  
A:Cross-references: GB:M20386  
R:Nilsson, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M.  
Cell 41, 719-726, 1985  
A:Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro-  
A:Reference number: A00643; MUID:85228222; PMID:2988784  
A:Accession: A00643  
A:Molecule type: mRNA  
A:Residues: 585-1223 <NII>  
A:Cross-references: GB:M10066  
C:Genetics:  
A:Gene: erbB  
C:Superfamily: epidermal growth factor receptor; protein kinase homology  
C:Keywords: alternative splicing; ATP; autophosphorylation; glycoprotein; growth factor  
specific protein kinase  
F:1-30/Domain: signal sequence #status predicted <SIG>  
F:31-1223/Product: epidermal growth factor receptor #status predicted <MAT>  
F:31-654/Domain: extracellular #status predicted <EXT>  
F:81-307/Domain: EGF receptor extracellular domain repeat <E1>  
F:397-610/Domain: EGF receptor extracellular domain repeat <E2>  
F:653-677/Domain: transmembrane #status predicted <TM>  
F:678-1223/Domain: intracellular #status predicted <INT>  
F:719-984/Domain: protein kinase homology <KIN>  
F:727-735/Region: protein kinase ATP-binding motif  
F:136-202,280,361,370,422,575,580,615,635/Binding site: carbohydrate (Thr) (covalent) #  
F:192,650/Binding site: carbohydrate (Ser) (covalent) #status predicted  
F:687/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted  
F:754/Active site: Lys #status predicted  
F:1100,1183,1208/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #stat

Query Match 31.4%; Score 1592.5; DB 1; Length 1223;

Best Local Similarity 29.0%; Pred. No. 2e-80; Matches 375; Conservative 135; Mismatches 310; Indels 473; Gaps 25;

8 RMGLLALALPPGAA-----STQVCTGTMKRLRPASPEHMDLRHLGGCGVOVGNLE 61  
Db 13 RGAAYVLLLVALLGVALSAVEKKVCGCTNNKLTQLGHVEDHFTSLQRMNCEVLSNLE 72  
Qy 62 LTYLPTNASLSPLODIOEVGYVLLAHANOVROVPLQRLRIVNGTOLEFENYALAVLNDG 121  
Db 73 IYVEHNRLDTLKTIOEAGVYLLALNVDVPLENLIQINGVLYDMSPLAVLSNH 132  
Qy 122 PLNNTPTVPGASBGRLREQLSLTEITLKGVLQGNPOLCYQDTLTKDIFHKNQAL 181  
Db 133 -NKKT-----GLRELPMKRLSELINGVAKISNNPKLCMDTVLMDIIDTSRK-PL 182  
Qy 182 TLID-TNRSRACHPCSPKCGSKNCGESESSEDCOSLRTVCAGGCA-RCKGPLPTDCCHQ 239  
Db 183 TVLDFAASNLSCPKCHPCTEDHCWAGAGNCOITLTKVCAQCCSGCKGKVPDCCNQ 242  
Qy 240 CAAGCGPKHSDCLAFHNSGICELHCPALVTYNTDTESMNPNEGRTFGASCVTAC 299  
Db 243 CAAGCGPRESDCLACRKRDAFCCKDPCPLVLYNPITYQMDVNEGKYSGATCVRRC 302  
Qy 300 PNYLSTDVSGCTVCPRLNNOEYTAEDGTQRECKSPCARVYCYGLGHEMLREAVATSA 359  
Db 303 PNYVYVTDHSGVSCVSCNTDYEV-EENGVRKCKKCDGLSCVNGGIGELGIIISINAT 361  
Qy 360 NIOEFAGCKKIFGLAFPESEGDPAASNTAPLOPQLOVFETLEITGYLISAMPDL 419  
Db 362 NIDSFNCKIKINGVSIIPVAFGLDAFTTLPDRKKLDVFTVKEISFGLIQAMPDA 421  
Qy 420 PDLVFNQLVQIRILHNGAVSLQGLIGISWLGSLRLDELGSGIALIHNTHLCEVHT 479  
Db 422 TDLYAFENLEIRGRTRKQGYSLAVVNLKIQSLRSLKEISDGDIALMKKNLCYADT 481  
Qy 480 VVMDQLFRPHQALHTARPREDECVBELACHOLCARHCNGPPTQCVNCSQPLRQGE 539  
Db 482 MMRSLSFATQSOXKTIQIRNRNNDCTADRHVCDPLCSVDVCGPGFHCFSRFSROKE 541  
Qy 540 CVEECRVLQGLPREVYNARHCLRPCHECOPQNG---SVTCFGEPPADQCAVCAHYKDPRC 596

Db 542 CVQNCNLLQGEPRFERNDSKCLPCHSECLVQNSTAVNTQCSGPGDHCKAKHFDGPHC 601  
Qy 597 VARCEGVNPDLSYMPIMKFPDEGACQPCPINCITSCYDLDDKCP----- 643  
Db 602 VKNCPAGVLGENDTL-VMKYADANAVCOLCHPCTGCGPGEGCPNGSKTPSIAGVY 660  
Qy 644 -----AEGRASPLRSQNE----- 656  
Db 661 GGLLCLVYVGLIGLYRRRHVYKRTLRLLQERLEVLTPSGEAPNQAHLRLIKETE 720  
Qy 657 ----- 656  
Db 721 FKKVYLVSGAGTGYKGLMIPBGEKVTIPVAKELREATSPRANKELIDPAVYMASVDN 780  
Qy 657 ----- 656  
Db 781 PHVCRLLGICLNSTVOLITQLMPYGLLDYIREHKDNIGSOYLAMWCQIAGMNYLEER 840  
Qy 657 -----DLS----- 659  
Db 841 RLVRDLAARNVLYKTPQHVKTTFGLAKLLGADEKEYHAEGKVPYIKWMALESILHRTY 900  
Qy 660 -----PA----- 661  
Db 901 THQSDVWSGYVWELMTFGSKPYDQIPASEISVLEKGERLPQPPICITDYVMIMVKW 960  
Qy 662 -----SPDSTFFYSRLDEDDMGDLV 682  
Db 961 MIDADSRPKFRELIAEFKSMARDPPRYLYIQDERMHLPSPDCKSRFLMLEEDMEDIV 1020  
Qy 683 DAEYLYVPOQGFPCPPAPAGAGMWHHRSSSTNSGGDLTLGLEPSEEARPSLT--- 739  
Db 1021 DDEYLYVPHQGF-----NSBST-----SRPTLSS 1046  
Qy 740 -APSEGASDVFDGDLGMAKGLQSLPTHPSPLORYSEDPVPLPSET--DGVAFL 795  
Db 1047 LSATSNKATNCID-----RNGCGHPRVEDSFYQRISSPTGNFLESIDDFL--- 1095  
Qy 796 TCSPOPEYVNPQDVRPQPSPREGLPAARPAATLERPKTLPKNGVYKDF----- 849  
Db 1096 -PAPEYVNO--LMPKRS-----TAMVQNOYNNISLTAISK 1129  
Qy 850 -----AFGGAENPEYLLTPQGAAPQPPPAFPAFNLYWDO-----DPP 892  
Db 1130 LPMDSRYQNSHSTAVDNPEYL-----NTQSPLAKTVPFESSPYWIOSGNHQINLNP 1181  
Qy 893 E-----RGAPPSTFKGPTAENPEYLGADV 918  
Db 1182 DYQDPLPNETKPNGLIKVPAENPEYLVANAP 1214

#### RESULT 7

A36325  
epidermal growth factor receptor - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 25-Jan-1991 #sequence\_revision 25-Jan-1991 #text\_change 10-Oct-1997  
C:Accession: A36325  
R:Petich, L.A.; Harris, J.; Raymond, V.W.; Blasband, A.; Lee, D.C.; Barp, H.S.  
Mol. Cell. Biol. 10, 2973-2982, 1990  
A:Title: A truncated, secreted form of the epidermal growth factor receptor is encode  
A:Reference number: A36325; MUID:90258888; PMID:2342466  
A:Accession: A36325  
A:Status: Preliminary  
A:Molecule type: mRNA  
A:Residues: 1-644 <PEP>  
A:Cross-references: GB:M37394  
C:Superfamily: epidermal growth factor receptor; protein kinase homology  
C:Keywords: alternative splicing; ATP; growth factor receptor

Query Match 30.2%; Score 1536; DB 2; Length 644;  
Best Local Similarity 44.6%; Pred. No. 1.3e-77;  
Matches 287; Conservative 98; Mismatches 234; Indels 24; Gaps 8;

QY 3 LAALCRMGILLALLPPGA-ASTOVCTGDMKRLRLPASPEHLDMLRLHLYGCGVQVGNLE 61  
 D 15 LAALCAAG-----GALKEEKVQOGTSNRLTQGTGFEDHFLSLQRFNFNCEVVLGNLE 66  
 QY 62 LTYLPNLSLFLDIOEVQVYLIAHQVROVPLQRLRYRGTOLEFDNTALAVLNGD 121  
 D 67 ITTVQRNDSLFLKTIQIOAGVYVILALNTVERILENQIIRGNALYENTYLAVALSN-- 124  
 QY 122 PLNNTPTVYGASPGRLRLQRLSLTEILKGGVLIQNRPOLCYODTILMKDIFHKNNOLAL 181  
 D 125 -----YGNKTKGLRLPRLNRLQELLIGAVRSNNPILCMETIQRNDYI-QVFLISN 175  
 QY 182 TLIDTNRS-RACHPCSPMKSGRCMGESSEDCQSLRTFYCAGGA-RCKGPLPTDCHQ 239  
 D 176 MSMDVQRLHLCGPKCDPCSPMGSCWGRGEENQKLTILCAQOCGRRCGRSPSCCHNQ 235  
 QY 240 CAAGCTGKHSDCLACLFHNHSGICELCPALVTYNTDFEESMPREGRYTFGASCYAC 299  
 D 236 CAAGCTGREDCLVCHFRDEATCKDTPMLTNPPTYQMDVNPBEKYSFGATCVAKG 295  
 QY 300 PNYLSTDVGSCTLVCPRLHNOEVTAEADGTORCEKCSKPCARVCYGLGMEHLREVAVTSA 359  
 D 296 PRNYVTYTHGSCVACGSDYEEV-EDGVSKCKKCDGCRKVCNGIGIGERKDTLSIAT 354  
 QY 360 NIOEFAGCKITFGSLAPLESFDDGPASNTAPLOEQLOVEFTEITGYLYISAMPDL 419  
 D 355 NIKHKYCTAISGDHLIPLVAFKGSFTRPPLDRLEILKTYKEITGFLIIQAMPENW 414  
 QY 420 PDLSTFQNLQVIRGRILHNGAVSLTLOGLSIMGLRSLRELGSLALIHNTHLCPVHT 479  
 D 415 TDLHAFENLEIRGTRKOHGQPSLAVVGLNTSLGLRSLKELSDVDVILSGRNLCIANT 474  
 QY 480 VPMDOLFENPQALHTANPREDECVGEGSLACHOLCARGHCGPPTQVNCOSQPLRGQ 539  
 D 475 INMKLFETPNQKTKIMNRAEKOCKATNHNCPICSESGCMGPRTQCVSQNRSRRE 534  
 QY 540 CVEBCRVLIQGLPREYVNAHCLPCHPECOQONGSVTCFGEPAODCVCAHYKDPFCYAR 599  
 D 535 CVDKKNILEGGEREVENSECIOCHPECLPOTMNTCTGRGPDNCIKCAHYVDGPHCVKT 594  
 QY 600 CPSPGVKPLSYMPIMKFPDEGACQPCPINCTHSCVDLDDGKC 642  
 D 595 CPSPGIMENNTL-VMKFADANNVCHLCHANCTYGCAGPGLKGC 636  
 RESULT 8  
 A47253  
 epidermal growth factor receptor, HER4 - human  
 C:Species: Homo sapiens (man)  
 C:Date: 22-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 18-Jun-1999  
 C:Accession: A47253  
 R:Plowman, G.D.; Culouscou, J.M.; Whitney, G.S.; Green, J.M.; Carlton, G.W.; Foy, L.; Ne  
 Proc. Natl. Acad. Sci. U.S.A. 90, 1746-1750, 1993  
 A:Title: Ligand-specific activation of HER4/p180erbB4, a fourth member of the epidermal  
 A:Reference number: A47253; MUID:93189574; PMID:8383326  
 A:Accession: A47253  
 A:Status: preliminary; not compared with conceptual translation  
 A:Molecule type: nucleic acid  
 A:Residues: 1-1308 <PI>  
 A:Cross-References: GB:L07868; NID:g337359; PIDN:AAB59446.1; PID:g337360  
 A:Note: Sequence extracted from NCBI backbone (NCBI:P:126842)  
 C:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C:Keywords: ATP; growth factor receptor  
 F:716-981/Domain: protein kinase homology <KIN>  
 F:724-732/Region: protein kinase ATP-binding motif  
 Query Match 30.1%; Score 1528.5; DB 2; Length 1308;  
 Best Local Similarity 27.7%; Pred. No. 7.3e-77;  
 Matches 372; Conservative 132; Mismatches 338; Indels 499; Gaps 25;  
 QY 9 WGLLALLPPGA-----STVCTGDMKRLRLPASPEHLDMLRLHLYGCGVQVGNLE 64  
 D 15 LAALCAAG-----GALKEEKVQOGTSNRLTQGTGFEDHFLSLQRFNFNCEVVLGNLE 66

D 8 WWWSLVAAGTVOPSDQSVACAGTENKLSLSLEDOQRALRKYYENCEVVMGNLEITS 67  
 QY 65 LPTNASTLFLDIOEVQVYLIAHQVROVPLQRLRYRGTOLEFDNTALAVLNGDPLN 124  
 D 68 IEHRDLSFLASVREYTVYVALNQFRYLPLENRIIRGKLYEDRALALIFLNYRKG 127  
 QY 125 NTPPTVYGASPGRLRLQRLSLTEILKGGVLIQNRPOLCYODTILMKDIFHKNNOLALTLI 184  
 D 128 NF-----GLQELGKNLLEILNGSVYDQNKFLCYADTIHMODIVRPMPSNLTLY 178  
 QY 185 DTNRSRACHPCSPMKSGRCMGESSEDCQSLRTFYCAGG-ARCKGPLPTDCHQCAAG 243  
 D 179 STNGSSGGRCHKSGTG-RCWGPTEHNCQTLTRFYCAQCGRCYGPVSDCCHECAGG 237  
 QY 244 CTGPKHSDCLACLFHNHSGICELCPALVTYNTDFEESMPREGRYTFGASCYACAPNY 303  
 D 238 CSGPKDTCFACAMNFDNSGACVYTCQPFYVYVNPPTFQLEHFNKYYTGACVAKKCPINF 297  
 QY 304 LSTDVGSCTLVCPRLHNOEVTAEADGTORCEKCSKPCARVCYGLGMEHLREVAVTSAIOE 363  
 D 298 V-VDSGSCVACGSPSSKMEV-ENGIRKMKPCTDLCPRKACDGIQGLSLMSAQTVDSNIDK 355  
 QY 364 FAGCKITGSLAPLESFDDGPASNTAPLOEQLOVEFTEITGYLYISAMPDL 423  
 D 356 FINCKINGNLFLVTGIGHDPYNAIEAIDPEKLVFETVETGFLINQSPMPMTDES 415  
 QY 424 VFQNLQVIRGRILHNGAVSLTLOGLSIMGLRSLRELGSLALIHNTHLCPVHT 483  
 D 416 VFSNLYTIGRNLVSGSLILKQGITSLQFQSLKEISACINITYITDMSNLCYHTIWT 475  
 QY 484 QLFNPHQALHTANPREDECVGEGSLACHOLCARGHCGPPTQVNCOSQPLRGQ 543  
 D 476 TLFSTINQRIYIRDKRAENCTAEGMVCNHLSSGCMGPPTDCLSCRRSRGICIES 535  
 QY 544 CRVLQGLPREYVNAHCLPCHPECOQONGSVTCFGEPAODCVCAHYKDPFCYAR 602  
 D 536 CNLYDGERFERENGISVCECDPQCKMEDGLTCHGPEPDNCTYCSHKDPNCEKCPD 595  
 QY 603 GVKPDLSTYMPIMKFPDEGACQPCPINCTHSCVDLDDGKC-----PABOR----- 647  
 D 596 GLQGANSE--IFKXADPRECHPCPNCTOGCNGPTSHDCIYYPWTHGSTLPQNAHPTLI 653  
 QY 648 -----ASPLTSSO----- 654  
 D 654 AAGVIGLFLIYVIGLFEAVVYRRKSIKKRALRFLTELEVERPLETPGTAPOAOLRIL 713  
 QY 655 -----NEDLGP----- 660  
 D 714 KETELKRVKVLGSGAFGVYVYKIMVBEGETYKIPVALIILNETTGPKANVERFMDALIMA 773  
 QY 661 ----- 660  
 D 774 SMDHRLVRLGLVCLSPITQVLTQMLRPGCLLEYHNEKNINGSLMLNKCVOYAKGMV 833  
 QY 661 ----- 660  
 D 834 LEERLVRDLAARNVLKSPNHVKITDFGLARLLEDGEKEYNADGKMPITKMALECIH 893  
 QY 661 ----- 660  
 D 894 YRKFTGSDVMSYGVTLWELMTFGKPYDGIPTREIPDLLEKGERLPORPITCTIDVYVM 953  
 QY 661 -----ASPLDSTFRSLLEDDM 678  
 D 954 VKCMIIDADSRKRELAAEFSRMANRDPQRLVLYOGDDRMKLPSTNDSKFTQNLDEDL 1013  
 QY 679 GDLVDAEEYLVPOQGFPCDPAPGAGVYHHRHRSSTRSGGDLTLGLBSEBAPRS 737  
 D 1014 EDMDAEEYLVV-QAFNIPP-----LYTSRARIDSNRS-----EIGHSPPAVTPMSG 1061  
 QY 738 -----PLAP-SEBAGSDVVDGDLGKMAAGLOSLPHND 769  
 D 1062 NQFVYRDGFAAEOGVSVYRAPRTSTIPEAVVQAGTAEIFDDSCNCTLRKPVAPHVOE 1121

[illegible]

Db 238 PODDCCFACRHPNDGSCAVPRRCQRIYVYNNKLTLPDLERNPHIKYGGVCAVASCNHNV-V 296

Qy 307 DVGCTVLCPLHNOEVAEDGTQRCCKSPCANVCTGLGMEHLREYVAVTSANTIOEPAG 366

Db 297 DQTSVCRACPDDKKEVD-KNGLKKCEPCPGGICSPACEGTGG--SRQTVDDSSNIDGFVN 353

Qy 367 CKRTFGSLAFLPESTDGGRASNTAPLQPEQLQVEETLEITGLYLYISANPDSLPDLSVFQ 426

Db 354 CTKLIGNDLFDLTGNGSPWHRIPALDPEKINVRTRVETGLYLNISQWPHMNFVS 413

Qy 427 NLQVIRGLILNNGAVS-LTLOGLGISWLGSLRELSGSLALHHNTHLGFVHPVPDOL 485

Db 414 NLTTTGGSLNRRGFSLLIMKNLVNTSLGFSLSLEISAGRTYISANQOLCHHSNMTKV 473

Qy 486 FRNPQALLHTA-NRPEDECVEGEGIACHQLCARGCHGPGPTQCVNCSQFLRGQCEVEEC 544

Db 474 LRGPTEERLDIKHNRPRRDCAVEGKVCDPICSSGGCGPGPGQGLSCSRNYSRGVCVTHC 533

Qy 545 RYLGGLPREYVNAHRCLEPCHECOPONGSVTCFEPREADQVACAHYRDPPRCVACPSGV 604

Db 534 NFLNEPREFEAFHEACEFCHPECCPMESTATCNCSSGSDTCAQCAHFRDGPCHVSSCPHG 593

Qy 605 KPDLSYMPIKRPDEBEGACORCPINCHTSCVDLDDKGC 642

Db 594 LG--AKGPITYKPYQNECRPCHECTGCGKPELDC 629

RESULT 10

JC4387

epidermal growth factor receptor homolog precursor - rat

N:Alternate names: ErbB3 protein; HER3 protein

C:Species: Rattus norvegicus (Norway rat)

C:Date: 17-Jan-1996 #sequence\_revision 19-Apr-1996 #text\_change 13-Nov-1998

C:Accession: JC4387

R:Hellner, N.J.; Kim, H.H.; Greaves, C.H.; Sietke, S.L.; Kolland, J.G.

Gene 165, 279-284, 1995

A:Title: Cloning of the rat ErbB3 cDNA and characterization of the recombinant protein

A:Reference number: JC4387; MUID:96096535; PMID:8522190

A:Accession: JC4387

F:Molecule type: mRNA

A:Residues: 1-1339 <HEL>

A:Cross-references: GB:U29339; NID:915389; PID:915330

A:Experimental source: liver

A:Note: The authors translated the codon AAC for residue 369 as Thr and GTT for resid

C:Comment: This protein is a functional heregulin receptor that transduces signals to

C:Genetics:

A:Gene: ErbB3

C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase hom

C:Keywords: ATP; growth factor receptor; liver; phosphoprotein; transmembrane protein

F:1-19/Domain: signal sequence #status predicted <SIG>

F:1-1339/Product: epidermal growth factor homolog #status predicted <MAT>

F:640-659/Domain: transmembrane #status predicted <TM>

F:705-970/Domain: protein kinase homology <KIN>

F:713-721/Region: protein kinase ATP-binding motif

F:933,1051,1136,1194,1196,1219,1257,1259,1273,1286,1325/Binding site: phosphate (Tyr)

Query Match 26.6%; Score 1351.5; DB 2; Length 1339;

Best Local Similarity 42.2%; Pred. No. 4.5e-67;

Matches 269; Conservative 87; Mismatches 253; Indels 29; Gaps 13;

Qy 3 LAALCRMGILLALLPPGCA---STQVCTGTDMKRLPASPEPTHIDMLRHLYQCGQVQGN 59

Db 7 LQVLCT---FLSLASRGSEMGNSQAVCPGTLNGLSVTGDDADNQQYOTLYKLYKECVCEYVMGN 62

Qy 60 LELVTLPTNASTSLQDIOEQGVVLLAHANQVRQVPLQRLRYNGCTOLFEDNVALNLDN 119

Db 63 LEIVLTGHNADLSFLQWTRVTAIVLVAAMNEFSVLPNLRVYKGTQVYDGKRAIFEM-- 120

Qy 120 GDPINNTTPVYGASPGGIRLEQLRLSLTEILKGVLIQNRQPLQXODRIILKKDIFKHNOL 179

Db 121 ---LNYNT---NSSHALRQLKFTQLRLTEILISGVYIEKKDKLCHMDPIDRDIVRAR--- 170

Qy 180 ALTLIDINRSACHPGCSBMCKGSRCKGSESSDDCSLTRVYCAQGC-ARCKGPLPTDCC 238

Db	171	GAEIYKNNKNGACPRCHVECKG - RCMGRPRDDCQILFTITICAPQCGNCFGRPNQCCHD	229
QY	239	QCAAGCTGGRKHSDCIACIAPHNHSGICEIACHPALVYNTDFESMNPBGRYTFGASCYTA	298
Db	230	ECAGGCSGPDDIDCFACRRFNDGACVPRCEPIYKNTLFTQLEBNPHTKYQYGVCVAS	289
QY	239	CPYNTLSTDVSGCTIYCPYLNQNEVYAEADGTQRCCKSCSPCARVCTGGLMEHLREPRATYS	358
Db	290	CPHNIV -VDQFFCVACPRDKMEVD -KKGKMKCEPCGGIJCPRACGTSG -SRAYQTAD	345
QY	359	ANIQEFACKIKFGSLAFPLESFDDPASNTAPLQPELOVETELETGYLYISAWPDS	418
Db	346	SNIDCFVWCTKILGMLDELITGLANDPMHKIPALDPEKLVNFRTRYREITGYLNDISWPPH	405
QY	419	LPDLSFONLQYIRRIIHNAGYS -LTLOGIGISWIGRISLREIASSGALIIHHNTHLEFV	477
Db	406	MHNFSVFNLTITIGRSILXNRGFSLLIMKNLNTVSLGRSLKEIAGRYIYSANQOLCYH	465
QY	478	HTVPMDOLFERNHQAHLTA -NRPEDECVGEGLACHOLCANRHCWGPRTQCVNCSQFLR	536
Db	466	HSIIMTTRILRGSEERLDIKYDRPLGECIACBEKVCDPLCSSGCGCGRAPRGQCLSRANRS	525
QY	537	GOECYEEBCRVLOGLPREYVYNAHCLPCHPECOPONGSVTCFGRPADQCVACAHYKDPFC	596
Db	526	EGVCVTHGNFLQGEEREFVHEACSPCHPECLPMEGTSTYNGSGSDACARCAHFRDGPHC	585
QY	597	VARCSGKVPDLSYPIPKFPEDEGACQCPRLNCHSC	634
Db	586	VNSCPHGILG -AKGPYIKYPRDAQNECRPCHENCYQCG	621

RESULT 11  
 S06142  
 protein-tyrosine kinase (EC 2.7.1.112) mrk-y precursor - southern platyfish  
 N:Alternate names: epidermal growth factor receptor homolog; kinase-related transforming  
 C:Species: Xiphophorus maculatus (southern platyfish)  
 C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 18-Feb-2000  
 C:Accession: S06142; S13809  
 R:Witbrodt, J.; Adam, D.; Mallitschek, B.; Maeuener, W.; Raulf, F.; Telling, A.; Roberts  
 Nature 341, 415-421, 1989  
 A>Title: Novel putative receptor tyrosine kinase encoded by the melanoma-inducing Tu loc  
 A:Reference number: S06142; MUID:90015140; PMID:2797166  
 A:Accession: S06142  
 A:Molecule type: DNA  
 A:Residues: 1-1166 <WIT>  
 A:Cross-references: EMBL:X16891; NID:965290; PIDN:CAA34770.1; PID:965291  
 R:Adam, D.; Maeuener, W.; Scharlt, M.  
 Oncogene 6, 73-80, 1991  
 A>Title: Transcriptional activation of the melanoma inducing Xmrk oncogene in Xiphophorus  
 A:Reference number: S13807; MUID:91125882; PMID:1846957  
 A:Accession: S13809  
 A>Status: Preliminary; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 821-1025, 'N',1027-1098, 'A',1100-1166 <ADA>  
 A:Cross-references: EMBL:X56319; NID:965284; PIDN:CAA3763.1; PID:965285  
 C:Genetics:  
 A:Gene: mrk  
 A:Map position: y  
 A:Introns: 872/3; 898/1; 947/1; 979/3; 1025/3; 1056/1  
 C:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C:Keywords: ATP; growth factor receptor; phosphotransferase; transmembrane protein; tyros  
 E:1-25/Domain: signal sequence #status predicted <SIG>  
 E:26-1166/Product: kinase-related transforming protein (Tu) #status predicted <MAT>  
 E:707-972/Domain: protein kinase homology <KIN>  
 E:715-723/Region: protein kinase ATP-binding motif

Query Match	25.9%	Score 1313.5;	DB 1;	Length 1166;
Best Local Similarity	41.6%	Pred. No. 4.9e-65;		
Matches 268; Conservative	93;	Mismatches 259;	Indels 25;	Gaps 12;

[illegible]

```

RESULT 12
A42032
epidermal growth factor receptor - chicken
C:Species: Gallus gallus (chicken)
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 18-Jun-1999
C:Accession: A42032
R:Flückinger, T.W.; Maibue, N.J.; Kung, H.J.
Mol. Cell. Biol. 12, 883-893, 1992
A:Title: An alternatively processed mRNA from the avian c-erbB gene encodes a soluble
A:Reference number: A42032; MUID:92123214; PMID:1732751
A:Accession: A42032
A:Status: Preliminary
A:Molecule type: RNA
A:Residues: 1-527 <Full>
A:Cross-references: GB:MW7637; NID:g211737; PIDN:AAA40759.1; PID:g211738
A:Experimental source: liver
A:Note: sequence extracted from NCBI backbone (NCBIN:76892, NCBIPI:76893)
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; growth factor receptor

```

Query Match	22.7%	Score 1155;	DB 2;	Length 527;
Best Local Similarity	44.0%;	Pred. No. 1.2e-56;		
Matches 227; Conservative	87;	Mismatches 182;	Indels 20;	Gaps 8;
Qy 11 LLLALLPGAAST-----OVCTGTMKRLRSPASPHNDMLHLYOGCGVOAGNELFTYL 65		:	:	:
Db 20 LLLLILGRVALCSAVEEKKVCGGTNNKLLQLQGHVEDHFTSLDRMKNCEVVSINSEIIVY 79		:	:	:
Qy 66 PTNASTLSFDIDIOEVOGVYLLAHNOVROPLQRLRIVRGTQFEDENYALAVADNDSPLNN 125		:	:	:

Db 80 EHNRLDTFLKTIQEVAGVYLALANMVDYPLENIQIRGNVLXNSPALVALSNYH-MNK 138  
 Oy 126 TTPVTGASPGGLRELQSLSTEILKGVLIQORNPOLCYQDTILMKDIFHKNNQALALTLID 185  
 Db 139 TQ-----GLRELPMRLSELILINGVAKISNNPKLCNMNDVYLMNDIIDTSRK-PLTVLID 189  
 Oy 186 -TNRSRACHPCSPMKGSRGWGSESDCOSLTRPYCAGGCA-RCGRLPTDCHOCAG 243  
 Db 190 FASNLSSCPKHPNCTEDHGWAGEONQOTLKVCIAOOCSGRCGKVPDCCHQCAAG 249  
 Oy 244 CTGPKHSDCLACLFHNSHIGELHCPALVYNTDTFESMPNPRGRTYTGASCVTACPYNY 303  
 Db 250 CTGPRESDDLACRKFRDADATCKDPCPLVLYNPTTYQMDVNPREGKYSFGATVRCRPHNY 309  
 Oy 304 LSTDVSGCTLVCPILHNOEVTAEADGTQRCCKSPCARVYCYGLGMHLEVRRAVTSANTQ 363  
 Db 310 VYTBHSGSVRCNCTDYEV-ENGVRKCKKCDGLCSKVCNNGIGIELGLISINTNIDS 368  
 Oy 364 FAGCKKIGSLAFLEPSEDPASNTAPLOPELOLVFELEITGYLISAMPDLSPLDLS 423  
 Db 369 FKNCKIKGDVSLFPAVLGDAFTTLPDPRKLLVFTVKETISGFLIQAMPDADTLX 428  
 Oy 424 VFQNLQVIRGILHNGAVSLTLQGLISMLGLRLSRLSGLALIHNTLHLCFHTVPM 483  
 Db 429 AFENLEIRGRTRKHQGYSLAVNLKIQSLRLSLEISDDGIALMKKNLCYADTMNR 488  
 Oy 484 OLFRRPHQALLHTANRPDECVGEGSLACHOCARGH 519  
 Db 489 SLFATQSQKTRKLIQNRKNDK--SKSVCFPAFAKAH 522

## RESULT 13

A27131  
 A:Title: growth factor receptor - fruit fly (*Drosophila melanogaster*) (fragment)  
 C:Species: *Drosophila melanogaster*  
 C:Date: 19-Nov-1988 #sequence\_revision 19-Nov-1988 #text\_change 23-May-1997  
 C:Accession: A27131  
 R:Schaefer, E.D.; Segal, D.; Glazer, L.; Shilo, B.Z.  
 A:Title: Alternative 5' exons and tissue-specific expression of the *Drosophila* EGF recep  
 A:Reference number: A27131; MUID:87002474; PMID:3093080  
 A:Accession: A27131  
 A:Molecule type: mRNA  
 A:Residues: 1-843 <SCH>  
 C:Genetics:  
 A:Gene: FlyBase:Bgfr  
 A:Cross-references: FlyBase:FBgn0003731  
 C:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C:Keywords: ATP; growth factor receptor

Query Match 20.0%; Score 1014.5; DB 2; Length 843;  
 Best Local Similarity 33.1%; Pred. No. 1.le-48;  
 Matches 232; Conservative 92; Mismatches 280; Indels 97; Gaps 20;  
 Oy 24 QVCTSTDKRLRPAPEHDLMLRHLXGCGVQVGNLELTYP-T-NASLFDIOIEVOG 82  
 Db 51 KVCITKRLSLSPSKKEHNYRLNRRTYVDSNKLTLWRPNMDLSFLDINREYTG 110  
 Oy 83 YVLAHNOVOYPLQRLRIVRGTOLE-----EDNTALAVLNDGDPDLNTPVTGASPGGL 137  
 Db 111 YLILSHVAVKKVFKPLQIIRGRFLSLSEBEKALFY-----TYSGM 154  
 Oy 138 RELQRLSLTEILKGVLIQORNPOLCYQDTILMKDIFHKNNQALALTLIDTNSRACHPCSP 197  
 Db 155 YLLELPDLADVLNGVGFHNNYNLCHMRTIQMSELYVNGTAYVYVDTAPRECRPKCHE 214  
 Oy 198 MCKGSRGSESDCOSLTRPYCAGGCA--RCKGPLPTDCEOCACAGTGGKHSDDLAC 255  
 Db 215 SCTHG-CWGBQPKQAKSKSLTSCPOCAGRCYGPKRPECHLFGAGCGTGTOQDCTAC 273  
 Oy 236 LAFNHSIGLCHLCPALVYNTDTFESMPNPRGRTYTGASCVTACPYNYLSTDVSGCTLV 315  
 Db 274 KNFEFAVSKBECPRMKRYNPTTYVLEETNPBGKAYGATCVKECP-GHLLRNDNGACVASC 332

Oy 316 PLHNOEVTAEADGTQRCCKSPCARVYCYGLGMHLEVRRAVTSANTQIEPACGKTIIFSIA 375  
 Db 333 PDKKMDKGE-----CVPCNGPCPRTCPGVTLH-----AGNIDFRNCTVIDGNIR 379  
 Oy 376 FLPESEFDG--DPASNTA-----PLQPOLQVFELEITGYLISAMPDLSPLDVSFON 427  
 Db 380 ILDQTFSGQDYVANYTMRPIPLDPERREVFSTVKETITGLNTEGHHPGRNLSYTRN 439  
 Oy 428 LQVIRGILHNGAV-SLTLQGLISMLGLRLSRLSGLALIHNTLHLCFHTVPMDOLE 486  
 Db 440 LETINGRLQMESFAALAVKSSVLSLEMRNLKQSSSVYIQHNRDLCYNSINMPLAQ 499  
 Oy 487 RNPQALLHTANRPDECVGEGSLACHOCARHCCHGPRPTQCVNCSQFLRGQCEYECRV 546  
 Db 500 KEPEKVVWENLRADLEKNGTICSDCNEDEGCGAGTQDCTCKNFNFNGTCTADCGY 559  
 Oy 547 LQGLPREYVNAHRCPCRECPONGSVTCFGEPEADOCVACAHYKDPFVCARCPGVKP 606  
 Db 560 ISNAK--FDNRTCKICHPER-----TCNAGADHCOECYHVGQHCYSECPK--- 608  
 Oy 607 DLSTMPKRFPEDEGACOPCPINCTHSCVDDLDDKCPAQRASPLTSQNEDLGPASPL-- 664  
 Db 609 -----KYND-RGVCRECHATC-----DGC-----TGPRDTIGIAGCTTC 641  
 Oy 665 -----DSTFRSLLEDDMDGDLVDAEYLVPO-QGFPCP 697  
 Db 642 NLAIINDATYKRCCLKDKCPDGY-FWEYVHPQDGLSKP 681

## RESULT 14

E88257  
 A:Title: let-23 [imported] - *Caenorhabditis elegans*  
 C:Species: *Caenorhabditis elegans*  
 C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 24-Aug-2001  
 C:Accession: E88257  
 R:Anonymous, The C. elegans Sequencing Consortium.  
 A:Title: Genome sequence of the nematode *C. elegans*: a platform for investigating bio  
 A:Reference number: A75000; MUID:99069613; PMID:9851916  
 A>Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/projects/C-  
 A>Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999;  
 A:Accession: E88257  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1323 <STO>  
 A:Cross-references: GB:chr\_II; PIDN:CAA93882.1; PID:g3881523; GSPDB:GN00020  
 C:Genetics:  
 A:Gene: let-23  
 C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase hom

Query Match 14.1%; Score 718.5; DB 2; Length 1323;  
 Best Local Similarity 28.0%; Pred. No. 4.3e-32;  
 Matches 191; Conservative 107; Mismatches 256; Indels 129; Gaps 28;  
 Oy 25 VCTGDKMLRLPASPETHDLMLRHLXGCGVQVGNLELTYPN----- 68  
 Db 39 LCGSTTNGISRYGTGNI-LEDETMYRGCRRYGMELETTWIEANIKRKWRESTNSTVDPK 97  
 Oy 69 -----ASLSPLODIOEVOGYVLIHNOVOYRQVPLQRLRIVRGTOLEFEDNYALAVLNDGDP 122  
 Db 98 NEDSLPKSINFDNLEETIRGSLITIRANIOKISFRLKVIYGVDEFHON-ALYIKKNDK- 155  
 Oy 123 LNTTPVTGASPGGLRELQSLSTEILKGVLIQORNPOLCY-QDTILMKDIFHKNNQAL 181  
 Db 156 -----VHEVYMRRLVIRNGSVYIQDNPCKYIGKIDIMKELLYPD--VQ 199  
 Oy 182 TLIDTNSRACH-----PCSPMKGSRGWGSESDCOSLTRPYCAGGCAAC--KPL 231  
 Db 200 KYETNSHQHCYQNKSNMAKCHESC-NDKCMWGSDNDQRAYRVYVCPSQSCFYSNSTS 258  
 Oy 232 PTDCHHCOCAAGCCTPKHSDDLACLFHNSHIGLCHLCPALVYNTDTFESMPNPRGRTYF 291

```

Db      259 SYCCOSACAGGCTGCGHGPKNKCIACSKYELDGIETCPSKRIENHKRGRLVFNPDGRYON 318
      292 GASCVTACAPYNYL-SFVGSCTLYC-PLHNOEVAEDGTORCEKC-SKPCARVCYGLGME 348
      319 GNNVCKECPPELLIENDY--CVRHCSGDHNYDATKD--VRECEKCRSSSCPKITVDG-- 372
      349 HLEVRVAVTSANIOEFGACKRIEGLSLAFLESPDGPASNTAPLQPOLQVETLEITG 408
      373 HL-----TNETLKNLEGCQIDGHL-TIEHAF-----TYQGLKLEIVKIVSE 414
      409 YLTIASMPDLSPLDSVQNLQVIRGLILHNGAVSLT-OGIGISWLGRLSRLRELSGLAL 467
      415 YITIV--QGNFYDLKFLKNQIIIEGRKLHNVRMALAIYQDDLEELSLNLIKITGAVL 472
      468 IHNHTHLCFYHYVPMDOLE---RNPQALHTANRPDECEVGBGLACHOLCARGHCWGP 523
      473 IMKNHRLCTVSKIDWSSITTSKGDKNKPSLAIAENRDSKLCETEQRYCDKNKRGCGWK 532
      524 GPTQCVNCSQFLNGQECVECRVLOGLPREYVNAHCLPCHPRECOPONGSVTCFGEPAHQ 583
      533 EPEDECECTKTKMSVGTCEVCDT-KGFLRNQTSMK-CERCSPECE-----TCNGLELD 584
      584 CVACAH-----YKDPFCVARGCPGSKPDLSTMPIMKFPDEEGACQPCPINCTHSCVD 636
      585 CLTCRHNKTLVNSDFGNMECVHDC-----PVSHPTQKNVCEKCHPCY----- 628
      637 LDDKGPAPQORASPLTSQNEIDL 659
      629 --DNMC-----TGPDNLG 640

RESULT 15
S70712
protein-tyrosine kinase (EC 2.7.1.112) let-23 precursor - Caenorhabditis elegans
N:Alternate names: receptor tyrosine kinase let-23
C:Species: Caenorhabditis elegans
C:Date: 21-Apr-1997 #sequence_revise: 09-May-1997 #text_change 11-Jan-2002
C:Accession: S70712; S73101; S13422; T27682
R:Sakai, T.; Koga, M.; Ohshima, Y.
J. Mol. Biol. 256, 548-555, 1996
A:Title: Genomic structure and 5' regulatory regions of the let-23 gene in the nematode
A:Reference number: S70712; M0ID:9617760; PMID:8604137
A:Accession: S70712
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-1374 <SAK>
A:Cross-references: EMBL:D63426
A:Experimental source: strain N2
R:Koga, M.
submitted to the EMBL Data Library, July 1995
A:Reference number: S73101
A:Accession: S73101
A:Molecule type: DNA
A:Residues: 1-50, 'G', 52-1374 <KOG>
A:Cross-references: EMBL:D63466; NID:g1407562; PIDN:BA09729.1; PID:g1407563
A:Experimental source: strain N2
R:Arslan, R.Y.; Koga, M.; Mendel, J.E.; Ohshima, Y.; Sternberg, P.W.
Nature 348, 693-699, 1990
A:Title: The let-23 gene necessary for Caenorhabditis elegans vulval induction encodes a
A:Reference number: S13422; M0ID:9108019; PMID:1979659
A:Accession: S13422
A:Molecule type: mRNA
A:Residues: 52-1374 <ARO>
R:Thomas, K.
submitted to the EMBL Data Library, March 1996
A:Reference number: Z20404
A:Accession: T27682
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 52-1374 <WIL>
A:Cross-references: EMBL:Z70038; PIDN:CAA93882.1; GSPDB:GN00020; CESP:ZK1067.1
A:Experimental source: clone ZK1067

```

```

C:Genetics:
A:Gene: let-23; CESP:ZK1067.1
A:Map position: 2
A:Introns: 44/1: 51/1: 90/1: 112/3: 165/3: 217/1: 290/1: 379/1: 418/1: 448/2: 565/1:
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase hom
C:Keywords: ATP; phosphotransferase; transmembrane protein; tyrosine-specific protein
F:1-28/Domain: signal sequence #status predicted <Sig>
F:29-1374/Product: protein-tyrosine kinase let-23 #status predicted <MAT>
F:934-1199/Domain: protein kinase homology <KIN>
F:942-950/Region: protein kinase ATP-binding motif

Query Match      14.1%; Score 718.5; DB 2; Length 1374;
Best Local Similarity 28.0%; Pred. No. 4, 5e-32;
Matches 191; Conservative 107; Mismatches 256; Indels 129; Gaps 28;

      25 VCTGTDMKLLPASPETHDMLRLHYGCGVQVQGNLETLPLTN----- 68
      90 LCGSTTGIRYRGNGNI-LEDLFTMYRGCRVYGNLEITWIEANETIKMRESTNSTVDPK 148
      69 -----ASLSFLDIOEVQGVLIHNOVROVPLQRLIRYGTQLFEDNYALAVDNGDP 122
      149 NEDSPLKSNFEDLIERISLTIYRANIQKISPRIRAVIYGDVEFHDN-ALYTHNKDK- 206
      123 LNNTPVYTGASPGGLRELQRLSLEILKGVLIQRPOLCY-ODPIIMKDI FHNQDAL 181
      207 -----VHEVVMRELIVIRNGSVTTIQDNPKMICYIGDKIDMKELLYDPP--VQ 250
      182 TLIDNRSRACH-----PCSPKCKSRGCGSESEDCQSLTFRYCAGGARGC--KGPL 231
      251 KVEETNSHQHCYONGSKAMAKCHESC-NDKCGSGDNCCQRYRVCSCQCYNSTS 309
      232 PTDCHEOCAGAGCTGPKHSDCLACLFHNSGICELHCPALVTVNTDTEFESMPNEGRYTF 291
      310 SYECDSACLCGGCTGCGHGPKNKCIACSKYELDGIETCPSKRIENHKRGRLVFNPDGRYON 369
      292 GASCVTACAPYNYL-STVGSCTLYC-PLHNOEVAEDGTORCEKC-SKPCARVCYGLGME 348
      370 GNNVCKECPPELLIENDY--CVRHCSGDHNYDATKD--VRECEKCRSSSCPKITVDG-- 423
      349 HLEVRVAVTSANIOEFGACKRIEGLSLAFLESPDGPASNTAPLQPOLQVETLEITG 408
      424 HL-----TNETLKNLEGCQIDGHL-TIEHAF-----TYQGLKLEIVKIVSE 465
      409 YLTIASMPDLSPLDSVQNLQVIRGLILHNGAVSLT-OGIGISWLGRLSRLRELSGLAL 467
      466 YITIV--QGNFYDLKFLKNQIIIEGRKLHNVRMALAIYQDDLEELSLNLIKITGAVL 523
      468 IHNHTHLCFYHYVPMDOLE---RNPQALHTANRPDECEVGBGLACHOLCARGHCWGP 523
      524 IMKNHRLCTVSKIDWSSITTSKGDKNKPSLAIAENRDSKLCETEQRYCDKNKRGCGWK 583
      524 GPTQCVNCSQFLNGQECVECRVLOGLPREYVNAHCLPCHPRECOPONGSVTCFGEPAHQ 583
      584 EPEDECECTKTKMSVGTCEVCDT-KGFLRNQTSMK-CERCSPECE-----TCNGLELD 635
      584 CVACAH-----YKDPFCVARGCPGSKPDLSTMPIMKFPDEEGACQPCPINCTHSCVD 636
      CLTCRHNKTLVNSDFGNMECVHDC-----PVSHPTQKNVCEKCHPCY----- 679
      637 LDDKGPAPQORASPLTSQNEIDL 659
      680 --DNMC-----TGPDNLG 691

```

Search completed: January 13, 2003, 14:49:14  
 Job time : 33.5363 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 13, 2003, 14:44:35 ; Search time 42.2593 Seconds

(without alignments)  
4480.838 Million cell updates/sec

Title: US-09-854-356-6

Perfect score: 5078  
Sequence: 1 METALCRNGLLALLPPGA.....TEKGTPTAENPEYLGLDVPY 919

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SPREMBL\_21:\*  
2: sp.archaea:\*  
3: sp.bacteria:\*  
4: sp.fungi:\*  
5: sp.human:\*  
6: sp.invertebrate:\*  
7: sp.mammal:\*  
8: sp.mhc:\*  
9: sp.organella:\*  
10: sp.plant:\*  
11: sp.podent:\*  
12: sp.virus:\*  
13: sp.vertebrate:\*  
14: sp.unclassified:\*  
15: sp.virus:\*  
16: sp.bacteriapi:\*  
17: sp.archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4438	87.4	1259	6 O18735	O18735 canis fam1
2	1874.5	36.9	419	4 Q9UK79	Q9UK79 homo sapien
3	1631	32.1	1209	11 Q9GX70	Q9GX70 ratius norv
4	1609	31.7	1210	11 Q9EP98	Q9EP98 mus musculu
5	1533.5	30.2	655	11 Q9WVF5	Q9WVF5 mus musculu
6	1517.5	29.9	643	11 Q9ERV6	Q9ERV6 mus musculu
7	1357	26.7	1328	13 P79754	P79754 fugu rubrip
8	1353	26.6	1165	13 Q9YH40	Q9YH40 xiphophoru
9	1252.5	24.7	1137	13 Q9WF6	Q9WF6 gallus gall
10	1217	24.0	367	11 Q8R2X1	Q8R2X1 mus musculu
11	1175.5	23.1	412	4 Q8WVY0	Q8WVY0 homo sapien
12	1155	22.7	537	13 Q90836	Q90836 gallus gall
13	1032	20.3	1433	5 Q9B1H9	Q9B1H9 anophelis g
14	1028.5	20.3	478	11 Q9ES60	Q9ES60 ratius norv
15	971.5	19.1	599	13 Q9PSH2	Q9PSH2 gallus gall
16	901	17.7	165	4 Q14256	Q14256 homo sapien

17	887	17.5	176	11 Q923V5	Q923V5 ratius norv
18	734	14.5	331	4 Q9BUD7	Q9BUD7 homo sapien
19	723	14.2	149	6 Q9BG66	Q9BG66 oryctolagus
20	656	12.9	1368	5 Q23821	Q23821 caenorhabdi
21	621	12.2	1717	5 Q26566	Q26566 schistosoma
22	620.5	12.2	1193	5 Q9Y1X8	Q9Y1X8 ephydattia f
23	409.5	8.1	1362	13 Q9PVZ4	Q9PVZ4 xenopus lae
24	404.5	8.0	150	6 Q9BG64	Q9BG64 oryctolagus
25	395	7.8	151	6 Q9BG65	Q9BG65 oryctolagus
26	366.5	7.2	1368	13 Q8UW85	Q8UW85 bombyx mori
27	366.5	7.2	1472	5 Q9U5A8	Q9U5A8 schistosoma
28	363.5	7.2	366	5 Q26569	Q26569 schistosoma
29	362	7.1	1369	13 Q8UW86	Q8UW86 paratichthy
30	359	7.1	1358	13 Q73798	Q73798 xenopus lae
31	350.5	6.9	136	11 Q99PK1	Q99PK1 mus musculu
32	345.5	6.8	1418	13 Q93457	Q93457 scophthalmu
33	340.5	6.7	1671	5 Q9NJV5	Q9NJV5 biomphalari
34	339.5	6.7	410	11 Q63720	Q63720 ratius norv
35	331	6.5	334	5 Q26567	Q26567 schistosoma
36	331	6.5	342	5 Q26568	Q26568 schistosoma
37	328	6.5	149	11 Q60494	Q60494 cavia sp. e
38	326.5	6.4	1245	13 Q9YGH8	Q9YGH8 scophthalmu
39	325	6.4	1418	13 Q8UW83	Q8UW83 paratichthy
40	322	6.3	469	11 Q63721	Q63721 ratius norv
41	316	6.2	1371	11 Q9QVW4	Q9QVW4 ratius sp.
42	315	6.2	1412	13 Q8UW84	Q8UW84 paratichthy
43	307	6.0	2144	5 Q9VD94	Q9VD94 drosophila
44	302.5	6.0	1749	5 Q8T0W6	Q8T0W6 echinococcu
45	296	5.8	946	5 Q9VJ04	Q9VJ04 drosophila

#### ALIGNMENTS

RESULT 1  
ID O18735 PRELIMINARY; PRT; 1259 AA.  
AC O18735:  
DT 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Erbb-2.  
OS Canis familiaris (Dog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
OX NCBI\_Taxid=9615;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Yokota H.;  
RT "cDNA cloning of erbb-2 from canine mammary gland.";  
RL Submitted (CCF-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB008451; BAA23127.1; -.  
DR HSSP; P11362; 1FGK.  
DR InterPro; IPR002048; EF-hand.  
DR InterPro; IPR000494; EGFR\_Ldomain.  
DR InterPro; IPR000719; Euk.pkinase.  
DR InterPro; IPR002174; Furin-like.  
DR InterPro; IPR001245; Tyr.pkinase.  
DR InterPro; IPR004019; YLP\_motif.  
DR Pfam; PF00757; Furin-like; 1.  
DR Pfam; PF00069; pkinase; 1.  
DR Pfam; PF01030; Recep\_Ldomain; 2.  
DR Pfam; PF02757; YLP; 2.  
DR ProDom; PD000001; Euk\_pkinase; 1.  
DR SMART; SM00261; FY; 3.  
DR SMART; SM00219; TYRKc; 1.  
DR PROSITE; PS00018; EF\_HAND; UNKNOWN; 1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
KW ATP-binding; Transferase; Tyrosine-protein kinase.  
SQ SEQUENCE 1259 AA; 137989 MW; E37364D9C4ACDA6 CRC64;

Query Match 87.4%; Score 4438; DB 6; Length 1259;  
Best Local Similarity 66.4%; Pred. No. 0;  
Matches 837; Conservative 31; Mismatches 50; Indels 342; Gaps 3;

QY 1 MELAALCRNGLLALLPPGAASYOCTGDMKRLRPASPTHLDMRLHYOGCOVVOGNTL 60  
DB 1 MELAALCRNGLLALLPPGAAGTOCTGDMKRLRPASPTHLDMRLHYOGCOVVOGNTL 60  
QY 61 ELTYLPTNASLSFLDIOEVQGYVLIANNOVQVPLQRLRIYRGTOLEFEDNTALAVLNG 120  
DB 61 ELTYLPTNASLSFLDIOEVQGYVLIANNOVQVPLQRLRIYRGTOLEFEDNTALAVLNG 120  
QY 121 DPLNNTPYTGASPGGLRELQRLSLTEILKGVLIQRPOLCYODTILMKDIFHKNNOLA 180  
DB 121 DPLEGIPAPGAAGGLRELQRLSLTEILKGVLIQRPOLCHQDTILMKDVFHKNNOLA 180  
QY 181 LTLIDNRSRACHPCSPMKGSRKMGSESESDOSLTRVCAGGACARCKGRLPTDCHDEC 240  
DB 181 LTLIDNRSRACHPCSPMKGSRKMGSESDOSLTRVCAGGACARCKGRLPTDCHDEC 240  
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMPNPGRTFGASCYTACP 300  
DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMPNPGRTFGASCYTACP 300  
QY 301 YNLTSTVGSCTLVCPPLHNOEYTAEDGTQRCCKSPCARVCYGLGMEHLREVRAVTSAN 360  
DB 301 YNLTSTVGSCTLVCPPLHNOEYTAEDGTQRCCKSPCARVCYGLGMEHLREVRAVTSAN 360  
QY 361 IOEFACKKIFGSLAFPSFOGDPASNTAPLOPELOVFELEETGYLYTISAMPDLSL 420  
DB 361 IOEFACKKIFGSLAFPSFOGDPASNTAPLOPELOVFELEETGYLYTISAMPDLSL 420  
QY 421 DLSVFONLQVIRGLIHNGASTLTLOGIGISWGLRSLRELSGALIHNTLHCFVHY 480  
DB 421 NLSVFNQLVIRGRVLDGAYSTLTLOGIGISWGLRSLRELSGALIHNTLHCFVHY 480  
QY 481 PMDQLEFRNHQALLHTANRPEDECVGEGLAQHLCARGHCWGPPTQCVNCSQFLRGQRC 540  
DB 481 PMDQLEFRNHQALLHTANRPEDECVGEGLAQHLCARGHCWGPPTQCVNCSQFLRGQRC 540  
QY 541 VECRCVLOGLPREYVABRCHLPCHECOPONSVTGCEGADOCVCAHYKPPFCVANC 600  
DB 541 VECRCVLOGLPREYVABRCHLPCHECOPONSVTGCEGADOCVCAHYKPPFCVANC 600  
QY 601 PGGVFDLSYMPIMKPEDEGACQPCPINCSTHSCVDLDKGCPAEORASPLTS----- 653  
DB 601 PGGVFDLSYMPIMKPEDEGACQPCPINCSTHSCVDLDKGCPAEORASPLTS----- 653  
QY 653 PGGVFDLSYMPIMKPEDEGACQPCPINCSTHSCVDLDKGCPAEORASPLTS----- 653  
DB 653 PGGVFDLSYMPIMKPEDEGACQPCPINCSTHSCVDLDKGCPAEORASPLTS----- 653  
QY 654 ILLAVVVGVLGILIKRRRQKIRKYYMRRLQETELVEPLTPSGAMPNOAMRILKETEL 719  
DB 654 ILLAVVVGVLGILIKRRRQKIRKYYMRRLQETELVEPLTPSGAMPNOAMRILKETEL 719  
QY 720 RKKVVLGSGAFGVYKGIPIPDGENVKIPYALKVIRENTSPXANKELDEAVYAGVSGP 779  
DB 720 RKKVVLGSGAFGVYKGIPIPDGENVKIPYALKVIRENTSPXANKELDEAVYAGVSGP 779  
QY 780 YVSRLLGICLTSTVOLQVLTMPYGCGLDHYREHGRGLSQDILLNMCVOIAKGSYLEYDR 839  
DB 780 YVSRLLGICLTSTVOLQVLTMPYGCGLDHYREHGRGLSQDILLNMCVOIAKGSYLEYDR 839  
QY 840 LVHRDLAARNLVKSPNHVKTIDFGLARLLIDETEHADGKVPIKMMALBSIPRPRPT 899  
DB 840 LVHRDLAARNLVKSPNHVKTIDFGLARLLIDETEHADGKVPIKMMALBSIPRPRPT 899  
QY 900 HOSDVMSYGVYWEMLTFCAKPYDGIIPAREIDPLEKGRLEDPORPICTIDVTIMVKMM 959  
DB 900 HOSDVMSYGVYWEMLTFCAKPYDGIIPAREIDPLEKGRLEDPORPICTIDVTIMVKMM 959  
QY 960 IDSECRPRELVAEFSRWARDPQRFVYIQNEDLGPASLDSTFYSLLEDDMDGLVYA 1019  
DB 960 IDSECRPRELVAEFSRWARDPQRFVYIQNEDLGPASLDSTFYSLLEDDMDGLVYA 1019  
QY 685 EELVYPQOGFCPPDPAFGAGVHHRHSSSTRSGGDLTLGLEPSEEBAPNSPLAPSEG 744

DB 1020 EELVYPQOGFCPEPPDPAAGTAHRRHSSSTRNGGELTLGLEPSEEBAPNSPLAPSEG 1079  
QY 745 AGSDVFDGDLGMAAGKGLSIPTHDPSPLQORSSEDTVPLPSTDDCVAPLTCSPQEVY 804  
DB 1080 AGSDVFDGDLGMAAGKGLSIPTHDPSPLQORSSEDTVPLPSTDDCVAPLTCSPQEVY 1139  
QY 805 NOPDVAPQPPSPREGPLPAPAPAGATLER-----PKTSLPGKNGVYKDVFAFGAVENDE 859  
DB 1140 NOPDVAPQPPSPREGPLPAPAPAGATLERPKTSLPGKNGVYKDVFAFGAVENDE 1199  
QY 860 YLTPGGAAPQPPHPPAFSPAFEDNLYYWDPPERGAPSTKGTPTAENPEYLGLDVPY 919  
DB 1200 YLTPGGAAPQPPHPPAFSPAFEDNLYYWDPPERGAPSTKGTPTAENPEYLGLDVPY 1259

RESULT 2  
ID Q9UK79 PRELIMINARY; PRT: 419 AA.  
AC Q9UK79;  
DT 01-MAY-2000 (Tremblrel. 13, Created)  
DT 01-MAR-2001 (Tremblrel. 16, last sequence update)  
DT 01-JUN-2002 (Tremblrel. 21, last annotation update)  
DE Herstatin.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxId=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=9415951; PubMed=10485918;  
RA Doherty J.K., Bond C., Jardim A., Adelman J.P., Clinton G.M.;  
RT "The HER-2/neu receptor tyrosine kinase gene encodes a secreted  
RT autoinhibitor";  
RT Proc. Natl. Acad. Sci. U.S.A. 96:10869-10874(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Doherty J.K., Clinton G.M., Adelman J.P., Evans A.J., Hennen W.D.;  
RT Submitted (SEP-2000) to the EMBL/Genbank/DBJ databases.  
DR EMBL: AF177761; AAD56009.2;  
DR InterPro: IPR00494; EGFR\_L\_domain.  
DR InterPro: IPR002174; Furin-like.  
DR Pfam: PF00757; Furin-like.  
DR Pfam: PF01030; Recep\_L\_domain; 1.  
DR SMART: SM00261; Fu; 1.  
SQ SEQUENCE 419 AA; 45472 MW; FEClBB347E2D030C CRC64;

Query Match 36.9%; Score 1874.5; DB 4; Length 419;  
Best Local Similarity 89.6%; Pred. No. 8; Indels 7; Gaps 2;  
Matches 353; Conservative 5; Mismatches 29; Indels 7; Gaps 2;

QY 1 MELAALCRNGLLALLPPGAASYOCTGDMKRLRPASPTHLDMRLHYOGCOVVOGNTL 60  
DB 1 MELAALCRNGLLALLPPGAAGTOCTGDMKRLRPASPTHLDMRLHYOGCOVVOGNTL 60  
QY 61 ELTYLPTNASLSFLDIOEVQGYVLIANNOVQVPLQRLRIYRGTOLEFEDNTALAVLNG 120  
DB 61 ELTYLPTNASLSFLDIOEVQGYVLIANNOVQVPLQRLRIYRGTOLEFEDNTALAVLNG 120  
QY 121 DPLNNTPYTGASPGGLRELQRLSLTEILKGVLIQRPOLCYODTILMKDIFHKNNOLA 180  
DB 121 DPLEGIPAPGAAGGLRELQRLSLTEILKGVLIQRPOLCHQDTILMKDVFHKNNOLA 180  
QY 181 LTLIDNRSRACHPCSPMKGSRKMGSESESDOSLTRVCAGGACARCKGRLPTDCHDEC 240  
DB 181 LTLIDNRSRACHPCSPMKGSRKMGSESESDOSLTRVCAGGACARCKGRLPTDCHDEC 240  
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMPNPGRTFGASCYTACP 300  
DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMPNPGRTFGASCYTACP 300  
QY 301 YNLTSTVGSCTLVCPPLHNOEYTAEDGTQRCCKSPCARVCYGLGMEHLREVRAVTSAN 360



```
Db 301 YNTLSTDVSGCTLYVCPHLNQEVTAEADGTQRCCKSCKCAR-----GTHSLPPRAAVPVP 355
QY 361 IQEPAGCKKIFGSLAFIPSEFDGDPASNTAPLOP 394
Db 356 LRMOFG--PAHPVLSFLRPSMDLVSAFYSLPLAP 387

RESULT 3
090X70 PRELIMINARY: PRT; 1209 AA.
AC 090X70;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DE 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Epidermal growth factor receptor.
GN EGFR.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FISHER; TISSUE=LIVER;
RX MEDLINE=90258888; PubMed=2342466;
RA Petch L.A., Harris J., Raymond V.W., Blasband A.J., Lee D.C.,
RA Earp H.S.;
RT "A truncated, secreted form of the epidermal growth factor receptor is
RT encoded by an alternatively spliced transcript in normal rat tissue.";
RL Mol. Cell. Biol. 10:2973-2982(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FISHER; TISSUE=LIVER;
RL Petch L.A.;
RL Submitted (NOV-1991) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC Gutteridge K., Dawson T.L., Earp H.S.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: M37394; AAF14008.1; .
DR HSSP: P11362; 1FGR.
DR InterPro: IPR000494; EGFR_L.domain.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR001245; Tyr_pkinase.
DR Pfam: PF00757; Furin-like; 1.
DR Pfam: PF01030; Recep_L.domain; 2.
DR PRINTS: PR00109; TYRKINASE.
DR ProDom: PD000001; Euk_pkinase; 1.
DR SMART: SM00261; FU; 3.
DR SMART: SM00219; TYRKC; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Receptor; transferase; Tyrosine-protein kinase.
SQ
SEQUENCE 1209 AA: 134891 MW; 96FEETFCCLB773 CRC64;

Query Match 32.1%; Score 1631; DB 11; Length 1209;
Best Local Similarity 29.9%; Pred. No. 3,3e-118;
Matches 363; Conservative 128; Mismatches 314; Indels 454; Gaps 26;

QY 3 LALALRWGLLALLPQGA-ASTOVCTGTDMKRLRASPETHLMDLRHLXGCGVQVGNLE 61
DB 15 LALALCAAG-----GALEEKRYCOGTSNRLTQLGTFEDHFLSLQRMFNNCVVLGNLE 66
QY 62 LFLPLPNASISFLQDIOEVGVYLLAHNQVROYRLRLRVGTQLEFEDYVALAVDNGD 121
DB 67 ITTVQRNYDLSFLKTIQEVAGVYLLALNVTVERPLENLQIRGNALYENTYALAVLSN-- 124
QY 122 PLNNTPTVTCSPGLREQLRSLTEILKGVLIQIRNPOLCYODTLIMKDIFFKNNQAL 181
```

```
Db 125 -----YGTNKTGLRELPNRNLQELIGAVRFSNNPILCNMTIOWRDIY-ODVPLSN 175
QY 182 TLIDITNRS-RACHPCSMGCKSGCWGESSDCCSLRTVAGGCA-RCKRPLTDDCHEQ 229
Db 176 MSMDYVRHNLGCKCPKCDPSCNGSCWGEGENCCKTKIKIAQCCSRRCGRSRSDCHQ 235
QY 240 CAAGCTGPKHSDCLACILHFHNSGICELHCPALTYNTDPRESMPNPGRTYFASCTAC 299
Db 236 CAAGCTGPRESDCLVCHRFDEATCKDTCPPMLYNTTYQMDVNPBGKTSFGATCYKCC 255
QY 300 PNYLSTDVSGCTLYVCPHLNQEVTAEADGTQRCCKSPCARVCYGLMEHLREAVTSA 359
Db 296 PRNVYVTDHSGCYRACGPDYEV-EEDGVSKCKKCDQPCAKYONGIGIGFKDTLSINAT 354
QY 360 NIDEPAGCKKIFGSLAFIPSEFDGDPASNTAPLOPELOLYFETLEETGLTISANPDSL 419
Db 355 NIKHFKYCTAISGDLHLPLPAFKGDSFTRPPIDPRELETIKYKETGFLTLQAMPENW 414
QY 420 PDLISVPONLQVIRGLIHNGAVSLTLOGIGISMLGSLRELSGLALIHNNHLCPVHT 479
Db 415 TDLHAFENLEITIGRTKQKHQOFSLAVVGLNITSLGRLSEISDGVYISGNRNLCYANT 474
QY 480 VPMDQLFRNPQALLHTANPREDECVGEGLACHQLCARGHCWGPPTQVCNCSQFLNGOE 539
Db 475 INMKKLFGFTNQKTKIMNNAEAKDCKATNHCNPDLCSSECGWGPEDVQSCQNSVSGRE 534
QY 540 CVCECRYLQGLPRELYNARHCLFCHPECPQNSVYTCFGEADQCVACAHYKDPFCVAV 599
Db 535 CVKCNKMLEBEPREFEVSSECIOCHPECLPOTNITCTGTGGRPNCKICAHYVGGPHCVKT 594
QY 600 CPSSGVKPDLSYMPIMKRPDEGACQPCPINCTGSCVVDLDKGC----- 642
Db 595 CPSSGIMGENNTL-VKRFADANNVCHLCNACTGCGACGKCGQOQPGKPIPSIATGIVG 633
QY 643 -----PAEQRASPLTSQNE----- 656
Db 654 GLLEIVVALGIGLPMRRQLVRRKRLRLQRELEVPETPSGEAPNQAHLRLKETEE 713
QY 657 ----- 656
Db 714 KIKIVLGSAGFTVYKGLMPEGEKVIPIVAKELREBASPANKELIDRAYVASVDNP 773
QY 657 ----- 656
Db 774 HVCRLIGICTSTVOLITQMLPYGCLLDYVREKDNIGSOYLLNKCVOYAKKNYLEDNR 833
QY 657 -----DLG----- 659
Db 834 LVHRDLAARNVLYKTPQHVKITPFGIAKLIGABEKEYHNAEGKVPKIMMALESILHRIY 893
QY 660 -----PA----- 661
Db 894 HOSDVMSYGVTVWELMTFGSKPYDGIPASEISSILEKGERLPPRICTIDVYMINVCMW 953
QY 662 -----SPLDSTFYRSLIEDDDMDIYD. 663
Db 954 IDADSRPKRELILFESKARDPORYLYVIGDERMHLPSSTDNEFRALMEEDMEDVVD 1013
QY 684 AEEYLVPQOGFCPPDPAPGAGGMVHNHRSSYRSGGDLTLLSESEEARSPPLAPSE 743
Db 1014 ADEYLILPQGGF-----NSPT-----STPILSL 1039
QY 744 GAGSDVFDGDLGCAKGLDSELTNHPSPLOKRSSEDPYVPLPSET--DGVAVPLTCSPOP 801
Db 1040 SANSN-----SSYVACINRNSCVRKBDADFQRYSSDPTSVLTLDNDIDDFL-----PVP 1089
QY 802 EYVNODPVRPQRPSPREGPLPARPAGATL-----BRPKTSLSGKGVYVDFAFAGAVE 856
Db 1090 EYTNQ-----SYV-KRPAGSVONPYTHNQPLHAPGRDLHTON--PHSNAYS 1133
QY 857 NPEYL-TPQGAAPORHPAPAFSPAFDNYLWDO-----DP-----PERGAPPS 899
Db 1134 NPEYLTWAO-----PTCLSSGFDSSALWYQKQSHQMSLDNPDYQDQDFFEKAKPKNG 1184
```

QY 900 TEKGTPAENPEYGLDVP 918  
||| ||||| ||| : |  
Db 1185 IFKG-PTAENAEYLRVAP 1202

## RESULT 4

Q9EP98 PRELIMINARY; PRT: 1210 AA.

DT 01-MAR-2001 (TREMblrel. 16, Created)

DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)

DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)

DE Epidermal growth factor receptor isoform 1.

GN EGFR.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN=C3H/101, 129/SvJ, AND 129/SVEVTAAC;

RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,

RA Schehl C.S., Pearshall R.S., Green P.J., Yee D., Lampland A.L.,

RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,

RA Mahle N.J.;

RT "Comparative genomic sequence analysis and isolation of human and

RT mouse alternative Egfr transcripts encoding truncated receptor

RT isoforms";

RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF275366; AAC28045.1; JOINED.

DR EMBL: AF275365; AAC28045.1; JOINED.

DR EMBL: AF275367; AAC24386.1; -.

DR HSSP: P11362; IFGK.

DR MGD: MGI:95294; Egfr.

DR InterPro: IPR000345; CytC\_heme\_bind.

DR InterPro: IPR000494; EGFR\_L\_domain.

DR InterPro: IPR000719; Euk\_Pkinase.

DR InterPro: IPR002174; Furin-like.

DR InterPro: IPR002290; Ser\_thr\_Pkinase.

DR Pfam: PF00757; Furin-like; 1.

DR Pfam: PF00065; Pkinase; 1.

DR Pfam: PF01030; Recep\_L\_domain; 2.

DR PRINTS: PR00109; TYRKINASE.

DR PRODOM: PD000001; Euk\_Pkinase; 1.

DR SMART: SM00261; FU; 5.

DR SMART: SM00230; S\_TKC; 1.

DR SMART: SM00219; TYTK; 1.

DR PROSITE: PS00190; CYTOCHROME\_C; UNKNOWN\_1.

DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.

DR PROSITE: PS0011; PROTEIN\_KINASE\_DOM; 1.

DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR; 1.

KW ATP-binding; Receptor; Transferase.

SQ SEQUENCE 1210 AA; 134840 MW; 62CD021C9DE32E18 CRC64;

Query Match 31.7%; Score 1609; DB 11; Length 1210;

Best Local Similarity 29.6%; Pred. No. 1.7e-116;

Matches 378; Conservative 130; Mismatches 314; Indels 454; Gaps 25;

QY 11 LLLALLPAGAA--STGYCTGTDMLRLPASPETHLMDLRHLHYGCGVVGQVNGLETLPTLN 68

Db 14 LITLALCAAGGALAEKKVCGGTSNRLTQGTGFEDHFLSLQRYNNNECVVILGNLEITYVQRN 73

QY ASLSFLDIOEVQGVYVLAHNOVROVPLQRLRYNGTOLFEDNVALAVLNDGDLNNTPT 128

Db 74 YDLSFLKTOIEVAGVYVLAHNTVERIPLENQIIRGNALYETVALALISN----- 124

QY VTGASPGGLRELOLRSLREIILKGGVLIQRIORPOLCYODITLWKDI----FHNNOALATLI 184

Db 125 -YGNRTGRLREIPRNLEIILGAVRFESNNPLTMDITQMRDLYQVNFMSMSDL---- 180

QY 185 DTRNRACHPCSPMCKGRCWGESESDQSLRTVCAGCA-RCKGPLPTDCHEQCAAG 243

Db 181 -QSHPSCKPCDPSCPNCSWGGEHCQKLIKICAQCSHRGRSPSCCHQCAAG 239

QY 244 CTGPKHSDCLACILPHNHSIGELHCPALVYNTDTFESMPNBEGRYTGACVYACPNY 303

Db 240 CTGPRESCLVQCKQKQDQATCKDTCPLMLYNPTTYQDVNPEGRYSFGATCVKCKPKNY 299

QY 304 LSTDVSGCTLVCPHLNNOEVTAEADGTORCEKSKPCARCYGLGMEHLREVAVTSANTOE 363

Db 300 VYTDHSGCYVRACGPYIEY-EDDGIRCKKCDGPKRKYCNIGIGIEFDITLSINATNKH 358

QY 364 FAGCKRIFGSLAFPESEFDGDPASNTAPLQVQVFEITETGYLYISAMPDPLS 423

Db 359 FKCYTAISGDHLILPVAEKDSFTRTPLDREILEIKYKEITGFLIQAMPDNTDLH 418

QY 424 VFONLOVIRGRILHNGAVALTLOGIGTSMIGIRSLREISGALIHNTHTLCFVHTVWD 483

Db 419 AFENLEIRGRKQKQSLVAVGGLNTLSLGRSLKETISDGVITSGRNLCVANTIMWK 478

QY 484 QLFNPHOALLHTANRPEDECVGEGLACHOLCARHCWGPPTQVCNCSQPLRGCEVEE 543

Db 479 KLFGPNQKTKIMNRAKCDKAVHVCNPLCSSGCGPREDVCQNVSRGCEVEK 538

QY 544 CRVLOGLPREYVNAHRCPCPECPQNGSVTCGPEADQVCAAHYNDPPCVARCSG 603

Db 539 CNLIEGEPRFEVNESECIOCHPECLPQAMNITCTGGRGPNQIQCAHYIDGHCVKTCBAG 598

QY 604 VKPDLSTYPIKPEDEBACOPCPINCHSGVDLDDKCPA----- 644

Db 599 IMGENNTL-VKRYADANNVCHLCHANCITYGCAGPQLQCEVWPSPKIPSTATGIVGLL 657

QY 645 -----BORASPLTSQNE----- 656

Db 658 FIVVVALGIGLFMRRHIVRKRLRLQERELVEPLPSPSEAPNOAHLRIKETEFAKI 717

QY 657 ----- 656

Db 718 KVLGSGAGTYVYKGLMPEGEKVIPLAIKELREATSPKANKELIDEAYVMASVDNPHVC 777

QY 657 ----- 656

Db 778 RLIGICLSTVOLITQMLPYPGCLDLYVEHKDNISQYLNMVQVIAKGMVLEDRRLVH 837

QY 657 -----DLG----- 659

Db 838 RDLAARNVLTVPQVHKITLDGLAKLGAEEKEYHAEGKVPYIKWMALESILHRYTHQS 897

QY 660 -----PA----- 661

Db 898 DVMSYGVTVWELMTFGSKRPYDGIIPASDISILEKGERLPORPCTIDVYIMVKNQWIDA 957

QY 662 -----SPDLSFYRSLLEDDMGDLVDAEE 686

Db 958 DSRPFRELILFEFSKMAADPORIYLYQDERMHLPSPLDSNFTYRALMEDEMEDVYVDAE 1017

QY 687 YLVPOGFECPDPAPAGAGVNHRRSSSTRSGGDLTLGLEPSEEARPSPLAPSEBAG 746

Db 1018 YLTPQGFF-----NSPST-----SKRPLLSLSAT 1043

QY 747 SDVFDGDLGMAKGLQSLPTHDPSPLOQYSEDPVLPSET--DGVAAPLTCSPQPEYV 804

Db 1044 SN-----NSTVACINRNGSCRVKEDAFLOKRYSSDPGAVTEDNIDAF-----PVEPY 1093  
 QY 805 NQDPVAPPPPPREGLPARAPAGATL-----ERPRTLSPGKNVYKDYAFGCAVEANE 859  
 Db 1094 NQ-----SVP-KRPAQSVQNPVYHNPQLPAPGRRLATON--PHSNANGNE 1137  
 QY 860 YL-TPGGGAAPQHPHPAPSPAFDNLYWDQ-----DP-----PERGAPSTTK 902  
 Db 1138 YLNTAQ-----PTCLSSGSPNPAIMIQKSHQMSLDNDPDYQDDFFPKETKPKNGIRK 1188  
 QY 903 GTPPAENPEYGLDVP 918  
 Db 1189 G-PTAENAEYLRVAP 1203  
 RESULT 5  
 ID Q9WVF5 PRELIMINARY; PRT; 655 AA.  
 AC Q9WVF5;  
 DT 01-NOV-1999 (TREMBLrel. 12, Created)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Epidermal growth factor receptor (Epidermal growth factor receptor isoform 3).  
 DE EGFR.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=LIVER;  
 RA Reiter J.L., Threadgill D.W., Danielson A.J., Schell C., Lampard A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R., Maille N.J.;  
 RT "Alternative transcripts from the Human and Mouse EGFR Genes Encode Carboxy-Terminal Truncated Receptors."  
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C3H/101, 129/SVJ, AND 129/SVEVYAC;  
 RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J., Sinclair C.S., Pearsall R.S., Green P.J., Yee D., Lampard A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D., Maille N.J.;  
 RT "Comparative genomic sequence analysis and isolation of human and mouse alternative Egfr transcripts encoding truncated receptor isoforms."  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=LIVER;  
 RA MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishi Y., Aizawa T., Hara A., Fukuishi Y., Konno H., Aachi J., Fukuda S., Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanka R., Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R., Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T., Fleischman W., Gaasterland T., Gissi C., King B., Kochia H., Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J., Schmitt L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T., Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G., Blake J., Botfield D., Boujunga N., Carninci P., de Bonaldo M.F., Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M., Gustincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H., Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P., Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N., Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F., Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L., Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlsuki S., Hayashizaki Y.;  
 RA "Functional annotation of a full-length mouse cDNA collection."  
 RT Nature 409:685-690(2001).

DR EMBL; AF124513; AAD44149.1; -  
 DR EMBL; AF275366; AAG28047.1; -  
 DR EMBL; AF275364; AAG28047.1; JOINED.  
 DR EMBL; AF275365; AAG28047.1; JOINED.  
 DR EMBL; AK004944; BAB23688.1; -  
 DR EMBL; AK004883; BAB23641.1; -  
 DR EMBL; AK004911; BAB23662.1; -  
 DR MGI:95294; Egfr.  
 DR InterPro: IPR000494; EGFR\_L\_domain.  
 DR InterPro: IPR002174; Furin-Like.  
 DR Pfam: PF00757; Furin-Like; 1.  
 DR Pfam: PF01030; Recep\_L\_domain; 2.  
 DR SMART: SM00261; FU; 3.  
 KW Receptor.  
 SQ SEQUENCE 655 AA; 72906 MW; 6B34063B1BC928CB CRC64;  
 Query Match 30.2%; Score 1533.5; DB 11; Length 655;  
 Best Local Similarity 44.4%; Pred. No. 5,8e-111;  
 Matches 284; Conservative 99; Mismatches 233; Indels 23; Gaps 7;  
 QY 11 LIALLPQAA--STQVCTGTDMKRLPASPETHLDMHLHYOGQVYOGNLELYLPTN 68  
 Db 14 LTLALCAAGALEERKVCQGTSMRLQLGTFEDHFLSLQRMYNCEVYLGNEITYVQHN 73  
 QY 69 ASLSFQDIOEYGVYLIHNOYRQVPLQRLIRVGTQLEEDNYALAVDNGDPLNNTTP 128  
 Db 74 YDLSFRTQIEVAGYVILINLTYERIPLENLQIRGNALYENTYALAIISN----- 124  
 QY 129 VTGASPGRLRELQRLSLTEILKGVLIQRPOLCYODTILMKDI----FKNNQALTLI 184  
 Db 125 -YGTNRTGLRELPMRLQELILGAVFNSNPILCNMDTIQWMDIYQNVMSMSDL---- 180  
 QY 185 DTRNSRACHPSCBCKSGKSGWGESSEDCSLTRIVYAGGCA-RCKPRLTDCCHQCAAG 243  
 Db 181 -QSHPSCKPCPCSPGSCGSGWGGEGNCKLTKIICAQCSHRRCRSPSDCHQCAAG 239  
 QY 244 CTGPKRSDDLACHFPHSGICELHCPALVTYNTDPESMPNDEGRYTFGASCVTACPYV 303  
 Db 240 CTGPRSDCLVCKRQDEATCKDTCPPLMYNPTTYIQMDYVNEGKTSFATCKKCPRYV 299  
 QY 304 LSTVGSCTLVCPRLHNOEYTAEDGTQCEKSCPKARCYGACMEHLREAVTSANIOE 363  
 Db 300 VYDHSQVRAAGPDYVE-EDGIRKCKCGCPKVCNGIGEGFNTLSINATNIKX 358  
 QY 364 FAGCKRTFGLPPEPSFGDPAANTAPLQPEQLOFETLEETGYLYISAMPDLSIDIS 423  
 Db 359 FYCTPAISGDLHLIPAFKGDSTFTPTPLDPRELEITKVTKEITGLLIQAMPDWTDLH 418  
 QY 424 VFQNLQVINGRLHNGVSLTLOGISWLGISRLRELGSLALHHNHTLQFVHTPMD 483  
 Db 419 AFENLEITIRKTKQHGQSLAVAGLITSLGRSLKEISDGVYISGNRLCYANTINMK 478  
 QY 484 QLEFRNHQALLTANRPDECEVGEGLACHQLCARGHCWPGPTQCNSQSFQGECEVE 543  
 Db 479 KLEFGPNQKTKIMNNAEKDKCAVNVNCPILSSSECCWPEPRDVCSCQNVARGRECVK 538  
 QY 544 CRYVLCGLPREYVNAHCLPCHPCEOPQNSVTCFGEADQYACAHYKDPVCVARGPSG 603  
 Db 539 CILTEEPREFEYNSCELCQHCPCILQAMNITCTGGRPNCTQCAHYIDGPRCVTCPSAG 598  
 QY 604 VKPDLSTYMPIMKPEEGACQPCPTNCTHSCVDLDDKGC 642  
 Db 599 IMGENNTL-VMKYADANNVCHLCHANCYGCAGPGIQQC 636  
 RESULT 6  
 ID Q9ERV6 PRELIMINARY; PRT; 643 AA.  
 AC Q9ERV6;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Epidermal growth factor receptor isoform 2.

```
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090.
RN [1]
RP SEQUENCE FROM N. A.
RC STRAIN=C3H/101, 129/SVJ, AND 129/SEVETAC;
RA Reiter J.L., Threadgill D.W., Green G.D., Strunk K.E., Danielson A.J.,
RA Sinclair C.S., Pearsall R.S., Yee D., Lampland A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Maibhe N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egfr transcripts encoding truncated receptor
RT isoforms."
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF215366; AAC28046.1; JOINED.
DR EMBL; AF215364; AAC28046.1; JOINED.
DR EMBL; AF215365; AAC28046.1; JOINED.
DR MGD; MGI:95294; Egfr.
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR002174; Furin-like.
DR Pfam; PF00757; Furin-like.
DR Pfam; PF01030; Recep_L-like; 1.
DR SMART; SM00261; FU; 4.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
KW Receptor.
SO SEQUENCE. 643 AA; 71476 MW; DEF22002C84911B1 CRC64;

Query Match 29.9%; Score 1517.5; DB 11; Length 643;
Best Local Similarity 44.7%; Pred. No. 1e-109;
Matches 281; Conservative 98; Mismatches 227; Indels 23; Gaps 7;

QY 11 LLLALPARGAA--STQVCTGDMKRLPASPETHLDMRLHLYQCGOVQGNLETLPLN 68
DB 14 LFLTCAAGALAEKKVGGTSTNRLTQGTGDFHLSLQRMVNNCEVYVGNLEITLYVRN 73
QY 69 ASLSFLDIOEYGVYLLAHNOVROVPLORLRIVGTQLFEDNVYLAVIDNDPLANTPT 128
DB 74 YDLSLTKTQIEVAGVYLLALMTVERIPLENQIINGNALYETATLALISN----- 124
QY 129 VTGASPGRLRELQSLTEILKGYLLIQNPOLCYQDTILMKDI---FHKNNQALTLI 184
DB 125 -YGTNRITGLRELPMRNLIQELIIGAVFSNNPLICMDPTLIQMDIYQVNFMSMDL--- 180
QY 185 DTNRARARPCSPMKSGRCWGESSEDCQSLTRYCAGGCA-RCGRLPTDCHEQCAAG 243
DB 181 -QSHSPCKPCDPPSCPNSSCWGGEGENCOKLTKIICAOQCSHRCGRSPSDCHNQCAAG 239
QY 244 CTGPKHSQCLACIHFNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACPNY 303
DB 240 CTGPRESCLVQKQODATCKDKTCPLMLYNPTTYQMDVNPFGKYSFGATGVKKCPKNY 299
QY 304 LSTDVGSCTLVCPRLNDEYTAEDGTQRCCKSPCARVYCYLGMEHLREAVRAYSANIOE 363
DB 300 VVTDHSGCVARCPDYEVF-EEDGIRKCKKCDGRCRKYVNGIGIEFDTLISANTNFKH 358
QY 364 FAGCKKIFGSLAFIPESDGPASNTAPLQPOLOVFETLEETIGYLYISAMPDPLDS 423
DB 359 FKCYTAISGDALILPVAKGDSFTETPLDPRELEITLTKVEITGFLIQAAMPDWTLH 418
QY 424 VFQNLQVIRGRILNHGAVALTLOGIGISMLGRSLRELSGALNLHNHTHLCFVHTVMD 483
DB 419 AFENLEIRGTQKHQGOISLAVNGLNTISLGRSLKETISDGVILISGRNNLCYATFWK 478
QY 484 QLFNRNQHALLTANRPEDCEVGEGLACHQCLARGHCWPGFTQCVNCSQFLRGCEVVE 543
DB 479 KLFGRPNKXTIMNNRAKCDKAVNHYCNPLCSSSGCWGPRRDVCSQNNSRGCEVEK 538
QY 544 CRVTLQGLREYVNAHCLPCRPDECPQNGSVYTCFGRPADQCVACAHYKDPREPCVACPSG 603
DB 539 CNILEGEPRFEVNSSECIOCHPECLPQAMNITTCGRGPDNCIQCAHYIDGPHCVTCPAG 598
```

```
QY 604 VKPDLSTYMPIMKPEDEGACQPCPINCH 632
DB 599 IMGENNTL-VMKYADANNVCHLCHANCY 626

RESULT 7
ID P79754 PRELIMINARY; PRT: 1328 AA.
AC P79754.
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Erib3.
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Takifugu.
OX NCBI_TaxID=31033;
RN [1]
RP SEQUENCE FROM N. A.
RX MEDLINE=99177347; PubMed=10077531;
RA Gelliner K., Brenner S.;
RT "Analysis of 148 kb of genomic DNA around the wnt1 locus of Fugu
RT rubripes."
RL Genome Res. 9:251-258(1999).
DR EMBL; AF056116; AAC34391.1; -.
DR HSSP; P11362; IFEK.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_kinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; Kinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TyrcKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
KW ATP-binding; Transferase.
SO SEQUENCE. 1328 AA; 148613 MW; A333039258B647E9 CRC64;

Query Match 26.7%; Score 1357; DB 13; Length 1328;
Best Local Similarity 37.7%; Pred. No. 8.5e-97;
Matches 317; Conservative 101; Mismatches 308; Indels 114; Gaps 26;

QY 9 WGLIALLP--GAASDQ---VCTGDMKRLPASPETHLDMRLHLYQCGOVQGNLEL 62
DB 4 WRLIMCVASRLRASQTOEAVCPGTONGLSSTGSGENQYNLNKDRYKGEIIMGNLEI 63
QY 63 TYLPTNASLSFLDIOEYGVYLLAHNOVROVPLORLRIVGTQLFEDNVYLAVIDNDP 122
DB 64 TOIESNMPFSFLKTRTYTGYVLLAHNFQEIPLQQLVIRGNSLYERRFALSVELN--- 120
QY 123 LNTPTVYASGAGRLQSLTEILKGYLLIQNPOLCYQDTILMKDI---FHKNNQALTL 182
DB 121 ---YKDG--PSGLNQLGMLNLTLEIDGVOIINNKLYRGVWYWDIT-RNNDADTE 173
QY 183 LIDTNRARARPCSPMKSGRCWGESSEDCQSLTRYCAGGC-ARCKRLPTDCHEQCA 241
DB 174 IQFNREBVCVH--KSC-GNVCWGGKQOQILRTYVACAPQNDRCFTSPSRDCHIECA 229
QY 242 AGCTGPKHSQCLACIHFNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACPY 301
DB 230 AGCKRLPTDCFACLFLFNDGACVAPQCPQTLIYNKQTFQMETNPNKAYQYSGISVSCPT 289
QY 302 NYLSTDVGSCTLVCPRLNDEYTAEDGTQRCCKSPCARVYCYLGMEHLREAVRAYSAN 360
DB 290 HFV-VDGSVCVSCPDKMEV--ERGSQRCQELSGLCPCGEGGAE--QROTVSSN 343
QY 361 IQEFAGCKKIFGSLAFIPESDGPASNTAPLQPOLOVFETLEETIGYLYISAMPDPLP 420
```

```

Db 344 IDSEFNCKTIGSLHFLVYGLIGDDDFKNVPELDAKKEVRYRTREITDLINISWPKELN 403
Qy 421 DLSVPONLOVYIRGRILLHNGAYSLTLQGLISWLGRLSRLSGALLIHHNTHLCVHTY 480
Db 404 DLSVFSLLTIQGRSLFKRFRSLVMKRIPTLSLGRSLRLEISDGSYISQNAHLCYHHYV 463
Qy 481 PMQLFRNPH-QALLHTANPREDECEGELACQLCARGHGCMRGYTCVNCNQIFLQGE 539
Db 464 NMQLRGRSVRANSLNSNKPMAECYADGVCDPLSDSCWGPDPQCLSCRNSRHGT 523
Qy 540 CVEECVILGLPREVYNAH-CLPCHPECOPONGSVTCFPEADQCACAHYDPECYA 598
Db 524 CVAGCHFNSGIRPREFAGLNCVACHPECKPQTGKASCTGPGADECMACKRFMDGYCMS 583
Qy 599 RCPGVKPDLSYPIWKFPDEBAGACOPPCINCTHSCVDLDDKCPAQRASPLTSGNEDL 658
Db 584 SCPAGVN-DGEKGLIRKFPNRREGHCPCQONCTQCGSPCLNDCLBAARLTISSGQITGI 642
Qy 659 GPASPLDSIFRYRSLLEDMDGDLVDAEYLVPOQGFPCPDPAAGAGNHNH---HRSS 714
Db 643 ALGVPRGLIFCLVL-----FFL-----GMLYHRGLAIRKR 673
Qy 715 STRSGGDLTLGLEPSEEBAPRSP-----LAPSE-----GAGSDVFPDGLMG- 757
Db 674 AMNR-----YIESESEFPLRGEGKGVHARILKIPDLKIRPLGSGVF-GYVSKGF 725
Qy 758 -----AAKGLSLPHTHDSPLQRYSE--DPTVPLPSETDGY-----VAPLTC 797
Db 726 WPEGETVRIVAIKTIQ-----DSSGRQTFEITDHLISMGLDHPYIVIRLGLIGCPGTC 780

```

```

RESULT 8
Q9YH40 PRELIMINARY: PRT: 1165 AA.
AC Q9YH40:
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Receptor tyrosine kinase, proto-oncogene.
GN XMRK.
OS Xiphophorus xiphidium.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorphia; Atherinomorpha;
OC Cyprinodontiformes; Poeciliidae; Xiphophorus.
OX NCBI_Taxid=8086;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RIO PURIFICATION:
RX MEDLINE=98241172; PubMed=9582016;
RA Dimitrijevic N., Winkler C., Wellbrock C., Gomez A., Duschl J.,
RA Altschmid J., Scharl M.;
RT "Activation of the Xmrk proto-oncogene of Xiphophorus by
RT overexpression and mutational alterations.";
RT Oncogene 16:1681-1690(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=RIO PURIFICATION:
RA Scharl M.;
RT Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: U53471; ADI0500.2; -.
DR HSSP: P1362; 1FGK.
DR InterPro: IPR000345; Cyrc_heme_bind.
DR InterPro: IPR000494; EGR_L_domain.
DR InterPro: IPR000719; Euk_kinase.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR001899; Gram_pos_anchor.
DR InterPro: IPR001245; Tyr_kinase.
DR Pfam: PF00757; Furin-like; 1.
DR Pfam: PF00069; kinase; 1.
DR Pfam: PF01030; Recep_L_domain; 2.
DR PRINTS: PR00109; TYRKINASE.

```

```

DR Prodom: PD000001; Euk_kinase; 1.
DR SMART: SM00261; RU: 3-261;
DR SMART: SM00219; TYRKC; 1.
DR PROSITE: PS00190; CYTOCHROME_C; UNKNOWN_1.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 2.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; kinase; transferase; tyrosine-protein kinase.
SQ SEQUENCE 1165 AA; 129614 MW; 7FEE38D871A7AE CRC64;

Query Match 26.6%; Score 1353; DB 13; Length 1165;
Best local similarity 42.1%; Pred. No. 1.5e-96;
Matches 273; Conservative 91; Mismatches 258; Indels 27; Gaps 12;

Qy 1 MELALCRWGLLALLPFG-AAST---OVCTGDMKRLRASPETHLMLRLHYOGCOV 55
Db 4 LELLLEL-----LILLLLISGRCCSTDPDRKVCQSTNQMTM---LDNHLYLKKMKMYSCNV 56
Qy 56 VQGNLELTLYLPNTASLFLQDIOEVGYLLIAHNOYRQVPLQRLRYRGTOLEFEDYATA 115
Db 57 VLENLEITTYQENQDLSFLQSIQIEVGYVLIANNEVSTIPVLVRLIRQONLYEGNFTLL 116
Qy 116 VLNDGDPPLNNTPTVTCASPGGLRELQRLSTLETILKGVLLIQRPOLCYQDTILMKDIFPK 175
Db 117 VMSNTYCK-NPSSP--DYVOGLKQDLQSLNITELISGVKYSNHPILCNVETIMMDIVDK 173
Qy 176 NNQALTLIDITNRSRACHPCSPCKSGKSGESSEDCQSILTRTVCAAGC-ARCKGRLPND 234
Db 174 TSNPTNLLIPHAERQCCQKDCPCVNGSCWAPRGHCQFTKLKLAECQCNRRCRGPKPID 233
Qy 235 CCHQCQACGTGPKKHSDDLACLHFNHSGICELHCPALVYVYNDTFESMNPBERITFGAS 294
Db 234 CCNEHCAGCTGPRATDCACRFNDGTCCKDPPKITYDIYSHQVYVNPNIKYTFGA 293
Qy 295 CVATCPYNTLSTPDVGSCTLYCPLHQEVTAEDTQCEKSKRCARVCYGLMEHREVR 354
Db 294 CYKEGCSNIVYIE-GACVSCSAGMLEVD-ENKRSCKRCQDVCYPCVCGIGIGISLNTI 351
Qy 355 AVTSANIQEFACGKRTFGSLAPLSEFDDDPASNTAPLOPEQVLETEITGYLYISA 414
Db 352 AVNSTIGSFSNCTKINGDIILNRNSFEEDDPHYKIGPMDPEHMLNTVKEITGYLYIWM 411
Qy 415 WPDLSLDLSVPONLOVYIRGRILLHNGAYS-LTLQGLISWLGRLSRLSGALLIHHNTH 473
Db 412 WPENMSTLSVFOULELIRRTTFSSRGFSFVVQVSHLQWLGLSSLKEVASAGNVILKNTPO 471
Qy 474 LCFVHTVPMQDLFRNPHQALLHTANPREDECEGELACQLCARGHGCMRGPTQVCNCSQ 533
Db 472 LKASITNMRRLFRSEDQSIETDART-----ENQTCNNESEDCWGPRTMVCYSLH 524
Qy 534 FLRGQCEVECRYLQGLPREVYNAHRLCPCHPECOPONGSVTCFPEADQCACAHYKDP 593
Db 525 VVRGRCVASCNMLQGEPEADQVGRVCHQCECLVQTLSTLCYGGPRANCSKCAHFQDG 584
Qy 594 PCVACPSGVKPDLSYPIWKFPDEBAGACOPPCINCTHSCVDLDDKGC 642
Db 585 PQCIPRCPHGLGDGDTL-TWKYADRMGQCQPCQONCTQCGSGPGLSGC 632

RESULT 9
Q9W6F6 PRELIMINARY: PRT: 1137 AA.
AC Q9W6F6:
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Receptor tyrosine kinase (Fragment).
GN ERBB4.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.

```

OX NCBI\_TaxId=9031;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=HINDRAIN;  
RX MEDLINE=99263203; PubMed=10328884;  
RA Dixon M., Lumsden A.;  
RT "Distribution of neuregulin-1 (nrg1) and erbB4 transcripts in  
embryonic chick hindbrain."  
RL Mol. Cell. Neurosci. 13:237-258(1999).  
DR EMBL: AF121963; AD31764.1; -  
DR HSSP: P11362; IFC.  
DR InterPro: IPR000494; EGFR\_Ldomain.  
DR InterPro: IPR000719; Euk\_Pkinase.  
DR InterPro: IPR002174; Furin-like.  
DR InterPro: IPR001368; TNFR\_c6.  
DR InterPro: IPR001245; Tyr\_kinase.  
DR InterPro: IPR004019; YLP\_motif.  
DR Pfam: PF00757; Furin-like; 1.  
DR Pfam: PF00069; Pkinase; 1.  
DR Pfam: PF01030; Recep\_L\_domain; 1.  
DR Pfam: PF02757; YLP; 2.  
DR PRINTS: PR00109; TYRKINASE.  
DR PRODOM: PD000001; Euk\_Pkinase; 1.  
DR SMART: SM00261; FU; 3.  
DR SMART: SM00219; TYRK; 1.  
DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE: PS00101; PROTEIN\_KINASE\_DOM; 1.  
DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR; 1.  
DR PROSITE: PS00652; TNFR\_NGFR\_1; UNKNOWN\_1.  
KW kinase; tyrosine-protein kinase.  
FT NON\_TER 1  
SQ SEQUENCE 1137 AA; 127927 MW; 4D616436F87D84F CRC64;  
Query Match 24.7%; Score 1252.5; DB 13; Length 1137;  
Best Local Similarity 26.1%; Pred. No. 9.6e-89;  
Matches 296; Conservative 121; Mismatches 314; Indels 403; Gaps 21;  
QY 161 LCYDFTILMKDIFHKNNALATLIDTNSRACHPCSPCKSGSCSESDQSLTRTVG 220  
DB 3 LCFADTIHMQDIVRNPMASNFTLVPTNGSSGGRCHKSCGT-RCMGPTFNHQQLTKTVG 61  
QY 221 AGGC-ARCKGLPTDCCHEQCAAGCTGPKHSDCLACLFHNHSGICELCPALVYINDTF 279  
DB 62 AEQCDGRCYGPGYVSDCCHEQCAAGCTGPKHSDCLACLFHNHSGICELCPALVYINDTF 121  
QY 280 ESMRPEGRYFEGASCYACPYNTLSTDVGSCTLVCPHNEVNAEDTOCEKSKRCA 339  
DB 122 QLENNHNAKTYTGAFCAKCPHNFY-VDSSCVRACPSSKMEV-ENGIKMKCPCTDTCR 179  
QY 340 RVCYGLGMEHLREYRAVYTSANIOEFAGCKIFGSLAFLPESFDGDPASNTAPLQEQLOV 399  
DB 180 KACOGIGTSLVSAQTVDSSNIDKINCTKINGNLIFLVTHGHPYTHIAINKEKINI 239  
QY 400 FETLEITGYLYISAMPDSDLVSYFONLOYIRGILHNGAVSLTGLGSLWSGLRSR 459  
DB 240 FQYREITGYLYNISWPMNMTDFRFYSNLVIGGRALYSGLSLILKQOGITSLQFOSLK 299  
QY 460 ELGSGALIHNTFLCEVHTVPMQDLERNPQALHTANRPEDECVGGLACHOICARGH 519  
DB 300 QISGNTIYTDNSNLCTHYTNWTSLESTPSQKTYIHNNKAENOTAGMVCNEICSSDG 359  
QY 520 CMGPGPTCVNCSQFLRQCEVCEGRVLAQLPREYVNAHCLPCHPECOR-ONGSVTFG 578  
DB 360 CMGPGPTCVNCSQFLRQCEVCEGRVLAQLPREYVNAHCLPCHPECOR-ONGSVTFG 578  
QY 579 PEADQCAACAHYKDPFVCANCPGSKVFDLSYMPWKFPEDEGACORPCINCHSCVDLD 638  
DB 420 PGRPHCTCFHFHKGDPNCEKCPDGLQGANF--IFKYADEDRCHPCNPTCTGCRGPA 477  
QY 639 DKGC-----PAQR----- 647  
DB 478 SHDCIYPTWSTLPQHAKRPLLAAGYIGLFIIVIMGLTFPAVYVRRKSIKKRALRRF 537

QY 648 -----ASPLTSQ----- 654  
DB 538 LETELVEPLTPSGAPNAQALRIKETELKRVKYLGSAGFGTVYKGIWVPGETKIPVA 597  
QY 655 -----NEDIGP----- 660  
DB 598 IKILETTGPRANVEFMDALIMASMDHPLVRLGLCPSPTIQLVTLMPHGLLDYVH 657  
QY 661 ----- 660  
DB 658 EHKDNISQILLNMCYOIAKGMVLEERRLVHRDLAARNLVKSPNHVKTDFGLARILE 717  
QY 661 ----- 660  
DB 718 GDEKEYNADGKMPDKWMALECIHYRKFTHSDVSYGVTIMEIMTEGKPYDGIPTREI 777  
QY 661 ----- 660  
DB 778 PDLEKGERLPDPICITIDVYVWVKWMIDADSRPKELAEFSRMARDPQRYLYIG 837  
QY 661 -----ASPLDSTYRSLLEDDMODLVDAEYLVPOGFCPPAPGAGVHRRHS 714  
DB 838 DDRKRLSPNSKFPQNLDEDELEDMDADEYLV-QAFTIPPTIYSTRIDSNNQF 896  
QY 715 STRSGGDLTLGLEPSEEDAPRS--PLAP-SEGASDVFDGLGAAKGLQSLPTHDS 771  
DB 897 VYRQGYAAEGGV-PMYRABGCIIPAPVAGATAEFEDTCNGTLRKQVATLAKEDS 955  
QY 772 PLQRYSEPTVPLPS-----EIDGYAPLTCSPQRPYVQDPVQRPQPPRGRGPLAA 824  
DB 956 STQRSADPTVFIPERYIRGELDEGVTPMRDKPKTDYILNVEENPVSRRKNGDLOA- 1014  
QY 825 RPAGATLERPK-TLSPGKNGVYKDV-----AFGAVENPEYLVPOGGAAPQPPPP 875  
DB 1015 -----VDNPEYHNAHPNQAPRADEYVNERLYMTFAITLENAYL-----KNLPE 1060  
QY 876 AFSPAEDNLVYWDQDPPERGA--PPSTFKGTPT-----AENDEYL 913  
DB 1061 KAKKAFNDPQVWNHSLPSTLQHPDYLQEXSTKYFYKQNGRIRPIVANEPEYL 1114  
RESULT 10  
Q8R2X1 ID 08R2X1 PRELIMINARY; PRT; 367 AA.  
AC 08R2X1;  
DT 01-JUN-2002 (Tremblrel. 21, Created)  
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)  
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
DE Hypothetical 40.2 kDa protein.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxId=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Strausberg R.;  
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL: BC027080; AAH27080.1; -  
KW Hypothetical protein.  
SQ SEQUENCE 367 AA; 40163 MW; 0BE03395F9E101B0 CRC64;  
Query Match 24.0%; Score 1217; DB 11; Length 367;  
Best Local Similarity 70.7%; Pred. No. 1.2e-86;  
Matches 241; Conservative 16; Mismatches 56; Indels 28; Gaps 5;  
QY 603 GVRDLSYMPIWKRPD--EAGACPCPICT-----HSCVDLDDKGCAPQASPLTS 653  
DB 31 GAKP-YDGIAPAREIPDLLEKGERLPQPICTIDYVIMVWKWMIDSECRP--RFRELYS 86  
QY 654 -----QNEDLGASPLDSTFYRSLLEDDMDLVDAEYLVPOGFCPPD 698  
DB 87 EFSRMARDPQRYLYIGLFIIVIMGLTFPAVYVRRKSIKKRALRRF 537

Qy	699	PAPGAGCVAHRRHNSSTSRGCGGLTGLLEPSEEARSPPLASEGAGSVPPGDLGMA	758
Db	147	PALGTGTAHRRHNSSTSRSGGGLTGLLEPSEEARSPPLASEGAGSVPPGDLAVG	206
Qy	759	AKGLQSLTHDPSPLOQRYSEDPYPLPSEINDGYAPLTCSPOEYVYVQDPVRQPSPRE	818
Db	207	TKGIQSLSPHLSPLQRYSEDPYPLPSEINDGYAPLTCSPOEYVYVQDPVRQSPRE	266
Qy	819	GPLPAAPAGATLERPKYTLSPGKNGVKDYAFEGGAVENPEYLTPOGGAAPQHPHPAFS	878
Db	267	GPPPIIPAGATLERPKYTLSPGKNGVKDYAFEGGAVENPEYLTAPRAGTASQHPHPAFS	326
Qy	879	PAFDNLYYMODPREBAPSPSTFGCGPTAANPEYLTGIDVAV	919
Db	327	PAFDNLYYMODNSSEGGPSPSTFGCTAANPEYLTGIDVAV	367

Query Match	Best Local Similarity	Matches 248;	Conservative 10;	Mismatches 45;	Indels 81;	Gaps 9;
QY 603 GVKPDLSTYMLPKFPD--EEGACGCPCLNCT-----HSCVDLDDKCGPARGASPLTS 653	23.1%; Score 1175.5; DB 4; Length 412;					
Db 31 GAKP-YDGIIPAREIDPLELEKGERLCPQPICTIDVYIMVKCMIDSECRP--RRELVS 86						
QY 654 -----QNEIDLGPASPLDSTFYRSLLEDDMDGDLVAEEIYVPOQGFPCPD 698						
Db 87 EFSRMARDPQRFVYIQNEDLDGPASPLDSTFYRSLLEDDMDGDLVAEEIYVPOQGFPCPD 146						
QY 699 PAPGAGGVNHRHRHSSSTRSGGGDLITLGLFSESEEAAPSPPLAPSGAGSDVFDGDLGMA 758						
Db 147 PAPGAGGVNHRHRHSSSTRSGGGDLITLGLFSESEEAAPSPPLAPSGAGSDVFDGDLGMA 206						
QY 759 AKGLQSLPTHDPSLQARSSEDPVPLPSETGQGYAFLTCSQPQEVVNPQVAPQPPSPRE 818						
Db 207 AKGLQSLPTHDPSLQARSSEDPVPLPSETGQGYAFLTCSQPQEVVNPQVAPQPPSPRE 266						
QY 819 GPLPAPRAGATLERPKTLSPGKNGVYKDVAFGCAVENPEYLPFGGAADP----- 871						

D <b>b</b>	267	GLPLAARPAAGATLERPTLS	SGKNGVKVDFAFAGGAENPE	LYLHPQGGALSP	LLPSA	326
Q <b>y</b>	872	-----HPPA---	ESPAEDNLYYND-QDPER	-----		894
D <b>b</b>	327	QPSTTSTGRTTHOSGGLH	PAAPSGLHQRQTOSTI	WVNTCCCEPBGV	KRSPDVS	SSGREG 366
Q <b>y</b>	895	-----GAPSTFKGT	PAEN 909			
D <b>b</b>	387	LTSAGIKRWEGPPTTS	RCOTCAHN 410			

RESULT 12	ID	Q90836	PRELIMINARY;	PRT:	527 AA.
AC	Q90836				
DT	01-NOV-1996	(TREMblrel. 01, Created)			
DT	01-NOV-1996	(TREMblrel. 01, Last sequence update)			
DT	01-JUN-2002	(TREMblrel. 21, Last annotation update)			
DE	EGF/TCF-alpha receptor precursor.				
CN	C-ERBB.				
OS	Gallus gallus (Chicken).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;				
CC	Gallus.				
OX	NCBI_TaxID=9031;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=92123214; PubMed=1732751;				
RA	Flickinger T.W., Malhale N.J., Kung H.-J.;				
RT	"An alternatively processed mRNA from the avian c-erbB gene encodes a				
RT	soluble, truncated form of the receptor that can block ligand-				
RT	dependent transformation."				
RL	Mol. Cell. Biol. 12:883-893(1992).				
DR	EMBL; M77637; AAA48759.1; -				
DR	InterPro; IPR000494; EGFR_L_domain.				
DR	InterPro; IPR002174; Furin-like.				
DR	Pfam; PF00757; Furin-like; 1.				
DR	Pfam; PF01030; Recep_L_domain; 2.				
DR	SMART; SMO0261; FU; 2.				
KM	Receptor; Signal.				
FT	SIGNAL	1	28	POTENTIAL.	
FT	CHAIN	29	527	EGF/TCF-ALPHA RECEPTOR.	
SO	SEQUENCE	527 AA;	58353 MM;	764564ABCC095298 CRC64;	
	Query Match	22.7%;	Score 1155;	DB 13;	Length 527;
	Best Local Similarity	44.0%;	Pred. No. 1.4e-81;		
	Matches 227;	Conservative 87;	Mismatches 182;	Indels 20;	Gaps
QY	11	LLALLPFGAASST-----QVCTGTDMKRLRSPETHIDMLRNHLKQSGGVGNLELYL	65		
DB	20	LLLLLLGVALCSANVEEKVCQGTNNKLTQLGHEVDHFTSLQRMNNEEVALNSLEIITY	79		
QY	66	PTNASLSTLDIOEYQGYVLLAHNOVRYPLQRLRIYVGTQLFEDNVALAVLDNGDPLNN	125		
DB	80	EHNRLDTFLKLTIOEAGVGLIALNMDVPIPLENGLIINGVALYDNSFALAVLSNYH-MNK	138		
QY	126	TTPVTGASPPGRLRETLLEIKGLVLIQNPOLCYODTILMKDIFHKNNQLATLTLD	185		
DB	139	TQ-----GRLPMLRLSEILLGVKISINPRLCMMDVLLMDLIDTSRK-PLTVLD	189		
QY	186	TNRSRACHPSCSPMKSGSRGSESEDDOSTLRYVCAGGCA-RCKGRLPTDCCHROCAAG	243		
DB	190	FASNLSSCPKHPNCTEDHCGAGSNGQTLTKVLCACQCSGRGKKRPSDCCNHQCAAG	249		
QY	244	CTGRHSDCLACLHPNHSIGELHCOPALVYVNTDTFESMPNDEGRYTFGASCVTAPYNY	303		
DB	250	CTGREGSCDLACRKFREDATCKDTCRPLVLYNPTTYQMDVNPDEGKYSFGATCYRECPHNY	309		
QY	304	LSTVGSTLVCPLRLHNEVTADEGROREKSKSCARVCYLSEMHLEFVRAVANSQOE	363		
DB	310	VYDHGSCVNSCMTDYEV-BENGRRCKKCKDGLGSCAKNSGIGELGLISTINNTNIDS	368		
QY	364	FAGCKKIFGSLAFIPESHSDGPASNTAPLQEOLOVFETLEIIGLYLISAMPDLPDLS	423		

Db 369 FNCKRTKNGVSVILPVALDPAFTKTLPLDPKRLDVRTYKEISGFLLIQAMPDNADLY 428  
QY 424 VFONLQYIRRIHNGVSVLTLQIGISMLGRLSRLGSLALIHNTLHCFVTVPW 483  
Db 429 AFENLEIIRRTKHOGGYSLAVYNLKTQSLGRLSKELISGDLAIMKNKLCYADTMNR 488  
QY 484 QLFERNPHQALLHTANRPEDECVGEGGLACRQLCARGH 519  
Db 489 SLFATOSQKRTKTIQNRKNDK--SKSVCFPAPAKAH 522

RESULT 13  
Q9BIH9 PRELIMINARY; PRT; 1433 AA.  
AC Q9BIH9;  
DT 01-JUN-2001 (TReMBLrel. 17, Created)  
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)  
DE Putative epidermal growth factor receptor (Fragment).  
OS Anopheles gambiae (African malaria mosquito).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;  
OC Anopheles.  
OX NCBI\_TaxID=7165;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-SUA:  
RA "lycett G.J.":  
RT "Cloning, expression and localisation of the Anopheles gambiae  
epidermal growth factor receptor."  
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
DR HSSP: P11362; IEGK.  
DR InterPro: IPR000345; Cytc\_heme\_bind.  
DR InterPro: IPR000494; EGFR\_L\_domain.  
DR InterPro: IPR000719; Euk\_pkinase.  
DR InterPro: IPR002174; Furin-like.  
DR InterPro: IPR002290; Ser\_thr\_pkinase.  
DR InterPro: IPR001245; Tyr\_pkinase.  
DR Pfam: PF00757; Furin-like. 1.  
DR Pfam: PF00069; Pkinase. 1.  
DR Pfam: PF01030; Recep\_L\_domain; 2.  
DR PRINTS: PR00109; TyRKINASE.  
DR PRODOM: PD000001; Euk\_pkinase; 1.  
DR SMART: SM00261; FU; 7.  
DR SMART: SM00220; S\_TKc; 1.  
DR SMART: SM00219; TYRKc; 1.  
DR PROSITE: PS00190; CYTOCHROME\_C; UNKNOWN\_4.  
DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE: PS00011; PROTEIN\_KINASE\_DOM; 1.  
DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR; 1.  
KW Receptor.  
FT NON\_TER.  
SQ SEQUENCE 1433 AA; 159585 MW; E3D9D88967724F07 CRC64;  
Query Match 20.3%; Score 1032; DB 5; Length 1433;  
Best local similarity 32.0%; Pred. No. 2e-71;  
Matches 251; Conservative 117; Mismatches 289; Indels 128; Gaps 26;

QY 200 KGRHWGESESDCOSLTRVACAGCA--RCKRLPTDCHEQCAAGCTGKHSDCIAC 257  
Db 165 EVG-CWEGAGNCCQRFSLKNCSPQCSOGRCFCGFPRECCILFCAGCCTGPTQSDIAC 223  
QY 258 FNSGICELCPALVYNTDTFESMPNBERGYTFGACVACAPYNTLSTVSGCTLVCP 317  
Db 224 FYDDGVCKQCEPPWQIYNPTNFMPEPBDKVAAGACVAKCP- EHLKNGACVRCPK 282  
QY 318 HNEFTEDEDTORCEKSCRCARCYGLGMEHLREVAVYASANOIEPAGCKIKFSLAF 377  
Db 283 GKMPQNSE-----CVPKGVCPRKTCPEGIVH-----SDNGKNKDCITIEGSEIL 329  
QY 378 PESFDGDPASNT-----APLQPEQLQVFELEITGYLYISAWDSLPLDSVFNQLQ 429  
Db 330 DQSPDFGOQVYTNFSFGPRYIKIDPRLLEVFSYVKEITGFINIQAHNPFTTYNFRNLE 389  
QY 430 VIRGRILHNGAY-SLTQIGIGISMLGRLSRLGSLALIHNTLHCFVTVPWDLFRN 488  
Db 390 VVGSRQLKENLIFASVYIVKTSLSLELSLKRNVSGSIVLLENSDLCFVEDIDMSEIKS 449  
QY 489 PHQALLHTANRPEDECVGEGGLACRQLCARGHCGPPTQCVNCSQFLRGQECVECRVLQ 548  
Db 450 SDHEVYVQKRNATKTECHEBEGMECSQCSKAGCMKGRPEQCLECKNVYKRGCLDSCK-- 506  
QY 549 GLPREY-VNARHCLPCHPECOQPONGSVTCFGEPADQCVACAHYKDPFCVYARCPGKY 607  
Db 507 SLPLVYSVDSKTCGDCHQCEKD-----FCYGPENDCNGCMNVGDFVCAECPT-TKHA 560  
QY 608 LSYPIWKFPDEBAGACQPCINCHSCVDDDDGCC--PAEQRASP-LTSQNDLGRPAS 663  
Db 561 MN-----GTC---INCHKTCV-----GCRGPRQTIAPDGCISCDKAIIGS--- 597  
QY 664 LDSTFFYSLLEDDMDGLVDAEYLVPOG-----FFC-----PDPAAGAGMVNHRH 711  
Db 598 -DAKIERCLMKDSCPRGKY-YSDYVLQEBGPLQOLSGKANCRCHPRCKKCTGYGFEQF 655  
QY 712 RSSSTRSGGDLTLGLPSESEAPRSLPSE-----GAGSD----- 748  
Db 656 COECTGYKKG-----ECCBDECPDIFYANEETRILCPHQECRGCHGLDDHNECRNLK 709  
QY 749 VFQSD 753  
Db 710 LFEED 714  
RESULT 14  
Q9ESEO PRELIMINARY; PRT; 478 AA.  
AC Q9ESEO;  
DT 01-MAR-2001 (TReMBLrel. 16, Created)  
DT 01-OCT-2001 (TReMBLrel. 18, Last sequence update)  
DE Putative epidermal growth factor receptor related protein.  
GN ERFP.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=GASTRO-DUODENAL MUCOUS;  
RA "yu Y., Moshier J.A., Majumdar A.P.N.":  
RT "Cloning of a novel EGFR-related peptide: A putative negative  
regulator of EGFR."  
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF187818; AAG17037.2; -  
DR InterPro: IPR000494; EGFR\_L\_domain.  
DR InterPro: IPR002174; Furin-like.  
DR Pfam: PF00757; Furin-like. 1.  
DR Pfam: PF01030; Recep\_L\_domain; 2.  
KW Receptor.  
SQ SEQUENCE 478 AA; 53233 MW; CF873A8376C519E5 CRC64;



Query Match	19.1%;	Score 971.5;	DB 13;	Length 599;
Best Local Similarity	30.2%;	Pred. No. 3.2e-67;		
Matches 222;	Conservative 76;	Mismatches 180;	Indels 257;	Gaps 11;

Search completed: January 13, 2003, 14:48:25  
Job time : 51.2593 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: January 13, 2003, 14:45:15 : Search time 12.232 Seconds  
(without alignments)  
1713.885 Million cell updates/sec

Title: US-09-854-356-7

Perfect score: 3954

Sequence: 1 MELAALCRWGLLALLPGGA.....GFCPPDPAPAGAGVHHRRH 712

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Issued Patents AA:\*  
2: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/PCITUS.COMB.pep:\*  
7: /cgn2\_6/ptodata/1/1aa/Backfilest.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3776	95.5	1255	2	US-08-625-101-2
2	3776	95.5	1255	2	US-08-356-786-2
3	3767	95.3	1255	1	US-08-467-083-68
4	3767	95.3	1255	1	US-08-414-417B-68
5	3767	95.3	1255	2	US-08-486-348A-68
6	3767	95.3	1255	2	US-08-468-545B-68
7	3767	95.3	1255	3	US-08-466-680B-68
8	3727	94.3	1255	2	US-08-484-438-8
9	3632	91.9	782	2	US-09-146-283-4
10	3632	91.9	782	3	US-08-579-823A-4
11	3632	91.9	782	4	US-09-344-155-4
12	3473	87.8	624	3	US-08-422-108-1
13	3473	87.8	624	4	US-08-422-734-1
14	1878	47.5	419	4	US-09-630-155-2
15	1534	38.8	1210	2	US-08-484-438-7
16	1534	38.8	1210	2	US-08-475-035-4
17	1533	38.8	644	1	US-08-336-708A-9
18	1458	36.9	911	2	US-08-484-438-10
19	1451	36.7	1058	2	US-08-484-438-4
20	1451	36.7	1308	2	US-08-484-438-2
21	1435	36.3	1342	2	US-07-978-895-4
22	1435	36.3	1342	2	US-08-484-438-9
23	1435	36.3	1342	2	US-08-473-119-4
24	1435	36.3	1342	2	US-08-475-352-4
25	1427.5	36.1	1343	6	US-09-570-454-2
26	1028.5	26.0	478	4	US-08-421-356-3
27	493	12.5	97	1	US-08-421-356-3

28	493	12.5	97	4	US-09-046-783-3	Sequence 3, Appl1
29	368	9.3	1382	2	US-08-737-715-2	Sequence 2, Appl1
30	368	9.3	1382	2	US-09-457-040B-7	Sequence 7, Appl1
31	350	8.9	580	1	US-08-414-417B-69	Sequence 69, Appl1
32	350	8.9	580	2	US-08-486-348A-69	Sequence 69, Appl1
33	350	8.9	580	2	US-08-468-545B-69	Sequence 69, Appl1
34	350	8.9	580	3	US-08-466-680B-69	Sequence 69, Appl1
35	313.5	7.9	1367	2	US-08-625-819-2	Sequence 2, Appl1
36	311.5	7.9	516	3	US-08-746-559A-4	Sequence 2, Appl1
37	311.5	7.9	1367	2	US-08-249-687C-2	Sequence 2, Appl1
38	311.5	7.9	1367	3	US-08-746-559A-2	Sequence 2, Appl1
39	303.5	7.7	1367	4	US-08-864-641B-18	Sequence 18, Appl1
40	295.5	7.5	486	3	US-08-746-559A-5	Sequence 5, Appl1
41	291	7.4	370	4	US-08-857-076-104	Sequence 104, App
42	267.5	6.8	1724	4	US-08-857-076-12	Sequence 12, Appl1
43	250	6.3	366	4	US-08-857-076-103	Sequence 103, App
44	223.5	5.7	383	4	US-08-857-076-105	Sequence 105, App
45	212	5.4	381	4	US-08-857-076-106	Sequence 106, App

## ALIGNMENTS

RESULT 1  
US-08-625-101-2  
Sequence 2, Application US/08625101  
Patent No. 5869445  
GENERAL INFORMATION:  
APPLICANT: Cheever, Martin A.  
TITLE OF INVENTION: COMPOUNDS FOR ELICITING OR ENHANCING IMMUNE  
TITLE OF INVENTION: REACTIVITY TO HER-2/neu PROTEIN FOR PREVENTION  
TITLE OF INVENTION: OR TREATMENT OF MALIGNANCIES IN WHICH THE HER-2/neu  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSER: SEED AND BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/625,101  
FILING DATE: 01-APR-1996  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Sharkey, Richard G.  
REGISTRATION NUMBER: 32,629  
REFERENCE/DOCKET NUMBER: 920010.448C7  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1255 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-625-101-2  
Query Match 95.5% Score 3776; DB 2; Length 1255;  
Best Local Similarity 67.9% Pred. No. 1.4e-317;  
Matches 712; Conservative 0; Mismatches 0; Indels 336; Gaps 1;  
QY 1 MELAALCRWGLLALLPGGAATGCTGCTDKMLRLPASPETHIDMLRYGCGQVQGNL 60  
DB 1 MELAALCRWGLLALLPGGAATGCTGCTDKMLRLPASPETHIDMLRYGCGQVQGNL 60

QY 61 ELTYLPTNASTSLFLODIOEVGYVLIANQVROVPLQRLRIVRGTOLEFEDNYALAVDNG 120  
 DB 61 ELTYLPTNASTSLFLODIOEVGYVLIANQVROVPLQRLRIVRGTOLEFEDNYALAVDNG 120  
 QY 121 DPLNNTPTVYGASPGGIRELQRLSLTEILKGVLIQRPOLCYODITLMDIFHKNNOLA 180  
 DB 121 DPLNNTPTVYGASPGGIRELQRLSLTEILKGVLIQRPOLCYODITLMDIFHKNNOLA 180  
 QY 181 LTLIDNRSRACHPCSPMCKSRGWGSSSDCOSLRTVACAGGARKGRLPDDCCHEOC 240  
 DB 181 LTLIDNRSRACHPCSPMCKSRGWGSSSDCOSLRTVACAGGARKGRLPDDCCHEOC 240  
 QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNREGRYTFGASCYTACP 300  
 DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNREGRYTFGASCYTACP 300  
 QY 301 YNYLSTDVGSCTIVCPPLHNOEVTAEADGTQRCCKSKPCARVCYGLGMEHLREVAVTAN 360  
 DB 301 YNYLSTDVGSCTIVCPPLHNOEVTAEADGTQRCCKSKPCARVCYGLGMEHLREVAVTAN 360  
 QY 361 IOEFAGCKKIFGSLAFPLESFDDGDPASNTAPLOEQLQVETLEITGYLYISAMPDPLP 420  
 DB 361 IOEFAGCKKIFGSLAFPLESFDDGDPASNTAPLOEQLQVETLEITGYLYISAMPDPLP 420  
 QY 421 DLSVFQNLQVIRGRILHNGAYSLTLOGLGISWLGSLRSLRELSGLALIHNTHLCEVHTV 480  
 DB 421 DLSVFQNLQVIRGRILHNGAYSLTLOGLGISWLGSLRSLRELSGLALIHNTHLCEVHTV 480  
 QY 481 PMOOLFNPHQALHTNANREDECVGSLACHOLCARGHMGWGPPOCVNCSOFLRGOEC 540  
 DB 481 PMOOLFNPHQALHTNANREDECVGSLACHOLCARGHMGWGPPOCVNCSOFLRGOEC 540  
 QY 541 VEECRVLOGLPREVYVNAHCLPCHECOPONGSVTCFGRPADQCAVAHKKDPFCVARC 600  
 DB 541 VEECRVLOGLPREVYVNAHCLPCHECOPONGSVTCFGRPADQCAVAHKKDPFCVARC 600  
 QY 601 PSGVKPDLSTYMPIMKPFDEGACQPCPINCTHSCVDLDDKGCRAQASPLTSISAVVG 660  
 DB 601 PSGVKPDLSTYMPIMKPFDEGACQPCPINCTHSCVDLDDKGCRAQASPLTSISAVVG 660  
 QY 654 ----- 653  
 DB 661 ILLVVLGVVFGILLIKRROOKIRKTYTMRRLQETELVEPLTPSGAMPNOAMRIKETEL 720  
 QY 654 ----- 653  
 DB 721 RKVKVILGSGAFGYVKGIWLPDGENVAKIPVAIKYLRNTPSKANKELIDEAYVAVGSP 780  
 QY 654 ----- 653  
 DB 781 YVSRLLGICLTSTVOLVTLMPYCGCLDHVRENKRLGSODLLMCMQIAKGMSTLEDDR 840  
 QY 654 ----- 653  
 DB 841 LVHRDLAARNVLYKSPNHVAKITDEGLARLLDIDETEXHADGKVPKIMMALESILRRFT 900  
 QY 654 ----- 653  
 DB 901 HQSDVMSYGYTWELMTFGAKPYDGIAPARELPDLEKGERLPORPCTIDVYIMTMVKCM 960  
 QY 654 ----- 653  
 DB 961 IDSECRFRPRELVSSESRMARDDPQRFVYIQNEDGPAASPDLSTYRSLLEDDDDMGDLVDA 1020  
 QY 685 EEYLVPOQGFCCPDPAFGAGMVAHHRH 712  
 DB 1021 EEYLVPOQGFCCPDPAFGAGMVAHHRH 1048

RESULT 2  
 US-08-356-786-2  
 ; Sequence 2, Application US/08356786

Patent No. 5877305  
 GENERAL INFORMATION:  
 APPLICANT: Huston, James S.  
 APPLICANT: Opperman, Hermann  
 APPLICANT: Houston, L. L.  
 APPLICANT: Ring, David B.  
 TITLE OF INVENTION: Biosynthetic Binding Protein for Cancer  
 NUMBER OF SEQUENCES: 16  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Edmund R. Pitcher, Testa, Hurwitz, & Thibault  
 STREET: Exchange Place, 53 State Street  
 CITY: Boston  
 STATE: Massachusetts  
 COUNTRY: USA  
 ZIP: 02109  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 SOFTWARE: PatentIn Release #1.0, version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/356,786  
 FILING DATE:  
 CLASSIFICATION: 424  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 07/831,967  
 FILING DATE: 06-FEB-1992  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Pitcher, Edmund R.  
 REGISTRATION NUMBER: 27,829  
 REFERENCE/DOCKET NUMBER: CRP-053  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617) 248-7000  
 TELEFAX: (617) 248-7100  
 INFORMATION FOR SEQ. ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1255 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-356-786-2  
 Query Match 95.5%; Score 3776; DB 2; Length 1255;  
 Best local similarity 67.9%; Pred. No. 1,4e-317;  
 Matches 712; Conservative 0; Mismatches 0; Indels 336; Gaps 1;  
 QY 1 MELAALCRWGLLALLPRGAASVQVCTGDMKRLRPSPEHMDMLRHLQGGCOVQGNL 60  
 DB 1 MELAALCRWGLLALLPRGAASVQVCTGDMKRLRPSPEHMDMLRHLQGGCOVQGNL 60  
 QY 61 ELTYLPTNASTSLFLODIOEVGYVLIANQVROVPLQRLRIVRGTOLEFEDNYALAVDNG 120  
 DB 61 ELTYLPTNASTSLFLODIOEVGYVLIANQVROVPLQRLRIVRGTOLEFEDNYALAVDNG 120  
 QY 121 DPLNNTPTVYGASPGGIRELQRLSLTEILKGVLIQRPOLCYODITLMDIFHKNNOLA 180  
 DB 121 DPLNNTPTVYGASPGGIRELQRLSLTEILKGVLIQRPOLCYODITLMDIFHKNNOLA 180  
 QY 181 LTLIDNRSRACHPCSPMCKSRGWGSSSDCOSLRTVACAGGARKGRLPDDCCHEOC 240  
 DB 181 LTLIDNRSRACHPCSPMCKSRGWGSSSDCOSLRTVACAGGARKGRLPDDCCHEOC 240  
 QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNREGRYTFGASCYTACP 300  
 DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNREGRYTFGASCYTACP 300  
 QY 301 YNYLSTDVGSCTIVCPPLHNOEVTAEADGTQRCCKSKPCARVCYGLGMEHLREVAVTAN 360  
 DB 301 YNYLSTDVGSCTIVCPPLHNOEVTAEADGTQRCCKSKPCARVCYGLGMEHLREVAVTAN 360  
 QY 361 IOEFAGCKKIFGSLAFPLESFDDGDPASNTAPLOEQLQVETLEITGYLYISAMPDPLP 420  
 DB 361 IOEFAGCKKIFGSLAFPLESFDDGDPASNTAPLOEQLQVETLEITGYLYISAMPDPLP 420

```

Db 361 IOEFACCKIFGSLAFPEPSFGDDPASNTAPLOPBOLOVEFTLEETGYLTISAMPDSL 420
Qy 421 DLSVFQNLQVINGRIILHNGAYSLTLQGLISWLGRLSRLSELGSLALIHNNHLCFVHTV 480
Db 421 DLSVFQNLQVINGRIILHNGAYSLTLQGLISWLGRLSRLSELGSLALIHNNHLCFVHTV 480
Qy 481 PMDOLFNRPHQALLHTANPREDECVGEGLAQHOLCARHCWGPGPTQCNCSQFLRGDGC 540
Db 481 PMDOLFNRPHQALLHTANPREDECVGEGLAQHOLCARHCWGPGPTQCNCSQFLRGDGC 540
Qy 541 VEECRVLOGLPREYVNAHRLCPHPECOPONSVCFCGEADQCVACAHYKPPFCVAC 600
Db 541 VEECRVLOGLPREYVNAHRLCPHPECOPONSVCFCGEADQCVACAHYKPPFCVAC 600
Qy 601 PSGVPRDLSTYMPIMKFPDEGACOPCPINCTHSCVDLDKGCAPABORASPLTSISAVYG 660
Db 601 PSGVPRDLSTYMPIMKFPDEGACOPCPINCTHSCVDLDKGCAPABORASPLTSISAVYG 660
Qy 654 ----- 653
Db 661 ILLVVLGVVFGILIKRQOKIRKTYMRLLOETELVEPLTPSGAMPNOQRILKETEL 720
Qy 654 ----- 653
Db 721 RKVYVLSGAFGTYYKGIWIPGENVKIPALKVLRENTSPRANKELIDEATVMAGVSP 780
Qy 654 ----- 653
Db 781 YVSRLLIGLITSTVOLYQMLPYGCLLDHVENRGLSQDILLNMCQIAKGMSTYLEDVR 840
Qy 654 ----- 653
Db 841 LVHRLAARNVLYKSPNHVKITDFGLARLLIDETEVHADGKVPILKMALESILRRFT 900
Qy 654 ----- 653
Db 901 HOSDWSYGVTVWELMTFGAKPYDGIAPAREIPDLLEKGRRLPQRPICITDYYMIMYKCM 960
Qy 654 ----- QNEDLGPSPLDSTFFRSLLEDDMDGLVDA 684
Db 961 IDSECRPRELVSSESRMARDPORFVYQJNEDLGPSPLDSTFFRSLLEDDMDGLVDA 1020
Qy 685 EEYLVPOOGFFCDDPARAGMWHHRH 712
Db 1021 EEYLVPOOGFFCDDPARAGMWHHRH 1048

RESULT 3
US-08-467-083-68
; Sequence 68, Application US/08467083
; Patent No. 5726023
; GENERAL INFORMATION:
; APPLICANT: Cheever, Martin A.
; TITLE OF INVENTION: IMMUNE REACTIVITY TO HER-2/NEU PROTEIN
; TITLE OF INVENTION: FOR DIAGNOSIS AND TREATMENT OF MALIGNANCIES IN WHICH THE
; NUMBER OF SEQUENCES: 68
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: US
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/467, 083
; FILING DATE: 06-JUN-1995

```

```

; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/414,417
; FILING DATE: 06-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Sharkey, Richard G.
; REGISTRATION NUMBER: 32,629
; REFERENCE/DOCKET NUMBER: 920010.448C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERY
; INFORMATION FOR SEQ ID NO: 68:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1255 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-467-083-68

Query Match 95.3%; Score 3767; DB 1; Length 1255;
Best Local Similarity 67.8%; Pred. No. 7,8e-317;
Matches 711; Conservative 0; Mismatches 1; Indels 336; Gaps 1;

Qy 1 MELAALCRWGLLALIPGAASVOYCTGDMKRLRSPSETHLMDRLHYOGCOVVOGML 60
Db 1 MELAALCRWGLLALIPGAASVOYCTGDMKRLRSPSETHLMDRLHYOGCOVVOGML 60
Qy 61 ELTYLPTNASLFLQDIOGVGVYLIANQVOPLORLIRYGTQLFEDNALAVLNG 120
Db 61 ELTYLPTNASLFLQDIOGVGVYLIANQVOPLORLIRYGTQLFEDNALAVLNG 120
Qy 121 DPLNTPYVTGASPGGLRELQRLSLTEILKGVLIQRPOLCYODTIIMKDFHKNQOLA 180
Db 121 DPLNTPYVTGASPGGLRELQRLSLTEILKGVLIQRPOLCYODTIIMKDFHKNQOLA 180
Qy 181 LFLIDTNRBACHPCSPMKGSRMGSESEDOSLRTVCAGGCAKRCPLTPDCHEQC 240
Db 181 LFLIDTNRBACHPCSPMKGSRMGSESEDOSLRTVCAGGCAKRCPLTPDCHEQC 240
Qy 241 AAGCTGPKHSDCLACLFHNSGICELCPALVTYNTDFESMPNREGRYTFGASCYTAC 300
Db 241 AAGCTGPKHSDCLACLFHNSGICELCPALVTYNTDFESMPNREGRYTFGASCYTAC 300
Qy 301 YNLTSDVSGCTLVCPPLHNOEYTAEDGTORCEKSPCARVCYGLAMEHLREYRAVTSAN 360
Db 301 YNLTSDVSGCTLVCPPLHNOEYTAEDGTORCEKSPCARVCYGLAMEHLREYRAVTSAN 360
Qy 361 IOEFACCKIFGSLAFPEPSFGDDPASNTAPLOPBOLOVEFTLEETGYLTISAMPDSL 420
Db 361 IOEFACCKIFGSLAFPEPSFGDDPASNTAPLOPBOLOVEFTLEETGYLTISAMPDSL 420
Qy 421 DLSVFQNLQVINGRIILHNGAYSLTLQGLISWLGRLSRLSELGSLALIHNNHLCFVHTV 480
Db 421 DLSVFQNLQVINGRIILHNGAYSLTLQGLISWLGRLSRLSELGSLALIHNNHLCFVHTV 480
Qy 481 PMDOLFNRPHQALLHTANPREDECVGEGLAQHOLCARHCWGPGPTQCNCSQFLRGDGC 540
Db 481 PMDOLFNRPHQALLHTANPREDECVGEGLAQHOLCARHCWGPGPTQCNCSQFLRGDGC 540
Qy 541 VEECRVLOGLPREYVNAHRLCPHPECOPONSVCFCGEADQCVACAHYKPPFCVAC 600
Db 541 VEECRVLOGLPREYVNAHRLCPHPECOPONSVCFCGEADQCVACAHYKPPFCVAC 600
Qy 601 PSGVPRDLSTYMPIMKFPDEGACOPCPINCTHSCVDLDKGCAPABORASPLTSISAVYG 660
Db 601 PSGVPRDLSTYMPIMKFPDEGACOPCPINCTHSCVDLDKGCAPABORASPLTSISAVYG 660
Qy 654 ----- 653
Db 661 ILLVVLGVVFGILIKRQOKIRKTYMRLLOETELVEPLTPSGAMPNOQRILKETEL 720
Qy 654 ----- 653

```

Db 721 RKVVLGSGAGFYVKGIMIPDGENVKIPVAKVIRENTSPKANKELIDEAYVAVAGVSP 780  
QY 654 ----- 653  
Db 781 YVSRLLGICLTSTVQVLTQMLPFGCLLDHVRENRRGLSGODLLWCMQIAKMSYLEDV 840  
QY 654 ----- 653  
Db 841 LVHRDLAARNLVKSPNHVKITDFGLARLDIDETEHADGKVPKMALESILRRFT 900  
QY 654 ----- 653  
Db 901 HQSDVWSYGVTVWELMTFGAKPYDGIAPARELPDLEKGERLPDPPICTIDVYIMVWKCM 960  
QY 654 ----- 653  
Db 961 IDSECRPRFRELVESEFSMARDPQRFVYIQNEDLGPSPLDSTFYRSLLDDEDDMDGLVDA 1020  
QY 685 EEYLVPOOGFPCPDPAAGAGMHHRR 712  
Db 1021 EEYLVPOOGFPCPDPAAGAGMHHRR 1048

RESULT 4  
US-08-414-417B-68  
; Sequence 68, Application US/08414417B  
; Patent No. 5801005  
; GENERAL INFORMATION:  
; APPLICANT: Cheever, Martin A.  
; TITLE OF INVENTION: IMMUNE REACTIVITY TO HER-2/neu PROTEIN  
; TITLE OF INVENTION: FOR DIAGNOSIS AND TREATMENT OF MALIGNANCIES IN WHICH THE  
; NUMBER OF INVENTIONS: HER-2/neu ONCOGENE IS ASSOCIATED  
; NUMBER OF SEQUENCES: 69  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Seed and Berry LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: US  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/414,417B  
; FILING DATE: 31-MAR-1995  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sharkey, Richard G.  
; REGISTRATION NUMBER: 32,629  
; REFERENCE/DOCKET NUMBER: 920010.448C2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 68:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1255 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; US-08-414-417B-68

Query Match 95.3%; Score 3767; DB 1; Length 1255;  
Best Local Similarity 67.8%; Pred. No. 7.8e-317;  
Matches 711; Conservative 0; Mismatches 1; Indels 336; Gaps 1;

QY 1 MELAALCRWGLLLALLPFGASTOVCTGDKMLRLPASPEHIMLRHLXGCGOVVGNL 60  
Db 1 MELAALCRWGLLLALLPFGASTOVCTGDKMLRLPASPEHIMLRHLXGCGOVVGNL 60  
QY 61 ELTYLPTNASLSFLDIOEVGVVLIANOVROYPLQRLIRVGTOLFEEDYALAVLDNG 120

Db 61 ELTYLPTNASLSFLDIOEVGVVLIANOVROYPLQRLIRVGTOLFEEDYALAVLDNG 120  
QY 121 DPLNNTPTVGTASPGGLRELOLRSTLETKGVLIQORNPOLCYODTIIKMDIFKNNOLA 180  
Db 121 DPLNNTPTVGTASPGGLRELOLRSTLETKGVLIQORNPOLCYODTIIKMDIFKNNOLA 180  
QY 181 LTLIDTNRSRACHPCSPMKSGRCWGSESSDQSLTRTVAGGACRCKGLPTDCHEQC 240  
Db 181 LTLIDTNRSRACHPCSPMKSGRCWGSESSDQSLTRTVAGGACRCKGLPTDCHEQC 240  
QY 241 AAGCTGPKHSDCLAFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVACP 300  
Db 241 AAGCTGPKHSDCLAFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVACP 300  
QY 301 YNVLSTDVGSCTLVCPPLHNOEVTAEADGTORCEKSKPCARCYGIGMEHLREVAVTSAN 360  
Db 301 YNVLSTDVGSCTLVCPPLHNOEVTAEADGTORCEKSKPCARCYGIGMEHLREVAVTSAN 360  
QY 361 IQEFAGCKKIRGSLAFLEPESFDGDPASNTAPLOPEQLOVEFTELEITGYLISAMPDLP 420  
Db 361 IQEFAGCKKIRGSLAFLEPESFDGDPASNTAPLOPEQLOVEFTELEITGYLISAMPDLP 420  
QY 421 DLSVFNQLVIRGRILHNGAVSLTLQIGISWLGRLSRLRELSGLALIHNTHLCPVHTV 480  
Db 421 DLSVFNQLVIRGRILHNGAVSLTLQIGISWLGRLSRLRELSGLALIHNTHLCPVHTV 480  
QY 481 PMDOLFNRPHQALHTANRPDECVGEGELACHOACARGHCGPPTOCVNCOSPLRQEC 540  
Db 481 PMDOLFNRPHQALHTANRPDECVGEGELACHOACARGHCGPPTOCVNCOSPLRQEC 540  
QY 541 VEECRVLQGLREYVNAHRCLPCHRECPONGSVYTCGPEADQCVACAHYDPPFCVARC 600  
Db 541 VEECRVLQGLREYVNAHRCLPCHRECPONGSVYTCGPEADQCVACAHYDPPFCVARC 600  
QY 601 PSQVNPDLISYPIKFPDEBACOPCPICTHSCVDDDKCPAEBQASPLTS----- 653  
Db 601 PSQVNPDLISYPIKFPDEBACOPCPICTHSCVDDDKCPAEBQASPLTS----- 653  
QY 654 ----- 653  
Db 654 ILVVVLGVVFGILLKRRQOKIRRYTMRLLQETELVEPLTPSGAMPNOAMRLKTEL 720  
QY 654 ----- 653  
Db 721 RKVVLGSGAGFYVKGIMIPDGENVKIPVAKVIRENTSPKANKELIDEAYVAVAGVSP 780  
QY 654 ----- 653  
Db 781 YVSRLLGICLTSTVQVLTQMLPFGCLLDHVRENRRGLSGODLLWCMQIAKMSYLEDV 840  
QY 654 ----- 653  
Db 841 LVHRDLAARNLVKSPNHVKITDFGLARLDIDETEHADGKVPKMALESILRRFT 900  
QY 654 ----- 653  
Db 901 HQSDVWSYGVTVWELMTFGAKPYDGIAPARELPDLEKGERLPDPPICTIDVYIMVWKCM 960  
QY 654 ----- 653  
Db 961 IDSECRPRFRELVESEFSMARDPQRFVYIQNEDLGPSPLDSTFYRSLLDDEDDMDGLVDA 1020  
QY 685 EEYLVPOOGFPCPDPAAGAGMHHRR 712  
Db 1021 EEYLVPOOGFPCPDPAAGAGMHHRR 1048

RESULT 5  
US-08-486-348A-68  
; Sequence 68, Application US/08486348A  
; Patent No. 5846538  
; GENERAL INFORMATION:



TELEFAX: (206) 682-6031  
 INFORMATION FOR SEQ ID NO: 68:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1255 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 US-08-468-545B-68

Query Match 95.3%; Score 3767; DB 2; Length 1255;  
 Best Local Similarity 67.8%; Pred. No. 7, 8e-317;  
 Matches 711; Conservative 0; Mismatches 1; Indels 336; Gaps 1;

```

OY 1 MELAALCRWGLLLALLPPGASTOVCTGTDKRLRLPASPTETHLMDLRHLYGCGVVOGNTL 60
DB 1 MELAALCRWGLLLALLPPGASTOVCTGTDKRLRLPASPTETHLMDLRHLYGCGVVOGNTL 60
OY 61 ELTYLPTNASLSFLDIOEVGYVLAHNOVROVPLRLRIVRGTOLEFEDNYALAVLDNG 120
DB 61 ELTYLPTNASLSFLDIOEVGYVLAHNOVROVPLRLRIVRGTOLEFEDNYALAVLDNG 120
OY 121 DPLNNTPTVVGASPGGLREIQLRSFTEILKGVLIOBNPOLCYODTILMKDIFKKNOLA 180
DB 121 DPLNNTPTVVGASPGGLREIQLRSFTEILKGVLIOBNPOLCYODTILMKDIFKKNOLA 180
OY 122 DPLNNTPTVVGASPGGLREIQLRSFTEILKGVLIOBNPOLCYODTILMKDIFKKNOLA 180
DB 122 DPLNNTPTVVGASPGGLREIQLRSFTEILKGVLIOBNPOLCYODTILMKDIFKKNOLA 180
OY 181 LTLIDTNRSRACHPCSPMKSGSRMGESSEDCQSLTRTVAGGACARCKGRLPTDCHEQC 240
DB 181 LTLIDTNRSRACHPCSPMKSGSRMGESSEDCQSLTRTVAGGACARCKGRLPTDCHEQC 240
OY 241 AAGCTGKHSDCLACLFHNSGICELCPALVYNTDTFESMPREBRYTFGASCYACR 300
DB 241 AAGCTGKHSDCLACLFHNSGICELCPALVYNTDTFESMPREBRYTFGASCYACR 300
OY 301 YNLTSTVSGCTVLCPLHNOEYTAEDGTORCEKSKPCARCYGLGMEHLREVAAYTSAN 360
DB 301 YNLTSTVSGCTVLCPLHNOEYTAEDGTORCEKSKPCARCYGLGMEHLREVAAYTSAN 360
OY 361 IOEAGACKIFGSLAFPLESPFDGDPASNTAPLOEOLQVEETLEITGYLYISAMPDLP 420
DB 361 IOEAGACKIFGSLAFPLESPFDGDPASNTAPLOEOLQVEETLEITGYLYISAMPDLP 420
OY 421 DLSYFOQLQYIRGILNNGAYSTLQGLGTSWGLRSLRELGSGLAIHNTHTLCFHTV 480
DB 421 DLSYFOQLQYIRGILNNGAYSTLQGLGTSWGLRSLRELGSGLAIHNTHTLCFHTV 480
OY 481 PMDOLFNPQALHTANREDECEVSGELACHOICARHGCHGPGPTOCVNSQFLRGQEC 540
DB 481 PMDOLFNPQALHTANREDECEVSGELACHOICARHGCHGPGPTOCVNSQFLRGQEC 540
OY 541 VEECRVLQGLPREYVNAHRLCPHRECOPONGSYTCFPGPEADQCVACAHYKDPFCVARC 600
DB 541 VEECRVLQGLPREYVNAHRLCPHRECOPONGSYTCFPGPEADQCVACAHYKDPFCVARC 600
OY 601 PSQVPRLSWPTKPFDEBAGACPRLNCTHSCVDLDDKCPAEQASPLTS----- 653
DB 601 PSQVPRLSWPTKPFDEBAGACPRLNCTHSCVDLDDKCPAEQASPLTS----- 653
OY 654 ----- 653
DB 654 ----- 653
OY 721 RKVVLGSGAFVYKGIWIPDGENKIPVAIKVLRENTSFKANKELIDEAYVAVAGVSP 780
DB 721 RKVVLGSGAFVYKGIWIPDGENKIPVAIKVLRENTSFKANKELIDEAYVAVAGVSP 780
OY 781 YVSRLLGICLTSTVOLTQMPYGCCLDHYVRENKRLGSDLLWMCQIAKMSYLEDVR 840
DB 781 YVSRLLGICLTSTVOLTQMPYGCCLDHYVRENKRLGSDLLWMCQIAKMSYLEDVR 840
OY 841 LVHHDLAARNVLYKSPHHVKITDGLARLLDIDETEVHADGKVPYIKMALESTILRRFT 900
DB 841 LVHHDLAARNVLYKSPHHVKITDGLARLLDIDETEVHADGKVPYIKMALESTILRRFT 900
OY 854 ----- 653
DB 854 ----- 653

```

```

DB 901 HSDVMSYGVTVWELMTFGAKPYDGLPAREIPDLLEKGERLPQPPICITIDVYIMKCMW 960
OY 654 -----ONEDLGPAASLDSTFYRSLIEDDMDGLVDA 684
DB 961 IDSECRARRELYSESRMARDPQRFVYIONEDLGPAASLDSTFYRSLIEDDMDGLVDA 1020
OY 685 EEXLVPOQGFCCPDPAAGAGVHHRRH 712
DB 1021 EEXLVPOQGFCCPDPAAGAGVHHRRH 1048

```

RESULT 7  
 US-08-466-680B-68  
 Sequence 68, Application US/08466680B  
 Patent No. 6075122

GENERAL INFORMATION:

APPLICANT: Cheever, Martin A.

TITLE OF INVENTION: IMMUNE REACTIVITY TO HER-2/neu PROTEIN

TITLE OF INVENTION: FOR DIAGNOSIS AND TREATMENT OF MALIGNANCIES IN WHICH THE

NUMBER OF SEQUENCES: 69

CORRESPONDENCE ADDRESS:

ADDRESSEE: Seed and Berry LLP

STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: US

ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/466, 680B

FILING DATE: 06-JUN-1995

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Sharky, Richard G.

REGISTRATION NUMBER: 32,629

REFERENCE/DOCKET NUMBER: 920010, 448C4

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 68:

SEQUENCE CHARACTERISTICS:

LENGTH: 1255 amino acids

TYPE: amino acid

TOPOLOGY: linear

US-08-466-680B-68

Query Match

Best Local Similarity 95.3%; Score 3767; DB 3; Length 1255;

Matches 711; Conservative 0; Mismatches 1; Indels 336; Gaps 1;

```

OY 1 MELAALCRWGLLLALLPPGASTOVCTGTDKRLRLPASPTETHLMDLRHLYGCGVVOGNTL 60
DB 1 MELAALCRWGLLLALLPPGASTOVCTGTDKRLRLPASPTETHLMDLRHLYGCGVVOGNTL 60
OY 61 ELTYLPTNASLSFLDIOEVGYVLAHNOVROVPLRLRIVRGTOLEFEDNYALAVLDNG 120
DB 61 ELTYLPTNASLSFLDIOEVGYVLAHNOVROVPLRLRIVRGTOLEFEDNYALAVLDNG 120
OY 121 DPLNNTPTVVGASPGGLREIQLRSFTEILKGVLIOBNPOLCYODTILMKDIFKKNOLA 180
DB 121 DPLNNTPTVVGASPGGLREIQLRSFTEILKGVLIOBNPOLCYODTILMKDIFKKNOLA 180
OY 122 DPLNNTPTVVGASPGGLREIQLRSFTEILKGVLIOBNPOLCYODTILMKDIFKKNOLA 180
DB 122 DPLNNTPTVVGASPGGLREIQLRSFTEILKGVLIOBNPOLCYODTILMKDIFKKNOLA 180
OY 181 LTLIDTNRSRACHPCSPMKSGSRMGESSEDCQSLTRTVAGGACARCKGRLPTDCHEQC 240
DB 181 LTLIDTNRSRACHPCSPMKSGSRMGESSEDCQSLTRTVAGGACARCKGRLPTDCHEQC 240

```





```

QY 421 DLSVFQNLQVIRGRILHNGAYSILTGLGISTWLGSLRLSRLSGSLALIHNTHLCEVHTV 480
DLSVFQNLQVIRGRILHNGAYSILTGLGISTWLGSLRLSRLSGSLALIHNTHLCEVHTV 480
DB 421 DLSVFQNLQVIRGRILHNGAYSILTGLGISTWLGSLRLSRLSGSLALIHNTHLCEVHTV 480
QY 481 PMDQLEFRNHQALHTANRDEDECVGEGLACHOLCARHGCHMGSPFQVCVCSQFLRGQEC 540
PMDQLEFRNHQALHTANRDEDECVGEGLACHOLCARHGCHMGSPFQVCVCSQFLRGQEC 540
DB 481 PMDQLEFRNHQALHTANRDEDECVGEGLACHOLCARHGCHMGSPFQVCVCSQFLRGQEC 540
QY 541 VEECRVLOGLPREYVNAHRLCPHRECOPQNSVTCGPEADQCVACAHKDDPFVCAR 600
VEECRVLOGLPREYVNAHRLCPHRECOPQNSVTCGPEADQCVACAHKDDPFVCAR 600
DB 541 VEECRVLOGLPREYVNAHRLCPHRECOPQNSVTCGPEADQCVACAHKDDPFVCAR 600
QY 601 PSGVKPDLSTYMPIMKFPDEGACQPCPINCTHSCVDLDDKGCRAEDRASPLTISAVNG 660
PSGVKPDLSYMPIMKFPDEGACQPCPINCTHSCVDLDDKGCRAEDRASPLTISAVNG 660
DB 601 PSGVKPDLSTYMPIMKFPDEGACQPCPINCTHSCVDLDDKGCRAEDRASPLTISAVNG 660
QY 654 ----- 653
DB 661 ILLVVLGVVFGILIRROOKIRKYMRLLOETELVEPLPSGAMPNQAOMRILETEL 720
ILLVVLGVVFGILIRROOKIRKYMRLLOETELVEPLPSGAMPNQAOMRILETEL 720
QY 654 ----- 653
DB 721 RKVKVLGSCAFGVYKGIWIPDGENKIPVAIVLRENTSPKANKELDEAYVMAGVGP 780
RKVKVLGSCAFGVYKGIWIPDGENKIPVAIVLRENTSPKANKELDEAYVMAGVGP 780
QY 654 ----- 653
DB 781 YVSRLLGICLTSTVOLVQLMPYGCILDHYENRGRLSQDLLNMCQIAKMSYLEDV 840
YVSRLLGICLTSTVOLVQLMPYGCILDHYENRGRLSQDLLNMCQIAKMSYLEDV 840
QY 654 ----- 653
DB 841 LVHRDLAARNVLYKSPNHVKITDFGLARLIDIDETEHADGKVPKIMMALESTLRRFT 900
LVHRDLAARNVLYKSPNHVKITDFGLARLIDIDETEHADGKVPKIMMALESTLRRFT 900
QY 654 ----- 653
DB 901 HQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLEKGERLPQPICTIDVYIMIMVCM 960
HQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLEKGERLPQPICTIDVYIMIMVCM 960
QY 654 ----- 653
DB 961 IDSECRREFRELVSERMAKRDQREVVIONEDLPASPLDSTFYRSLDDDDMDGLVDA 1020
IDSECRREFRELVSERMAKRDQREVVIONEDLPASPLDSTFYRSLDDDDMDGLVDA 1020
QY 685 EEXLVPOQGFPCPPAPAGAGMHHRR 712
EEXLVPOQGFPCPPAPAGAGMHHRR 712
DB 1021 EEXLVPOQGFPCPPAPAGAGMHHRR 1048
EEXLVPOQGFPCPPAPAGAGMHHRR 1048

RESULT 9
US-09-146-283-4
; Sequence 4, Application US/09146283
; Patent No. 5976546
; GENERAL INFORMATION:
; APPLICANT: Laus, Reiner
; APPLICANT: Ruegg, Curtis L.
; APPLICANT: Wu, Hongyu
; TITLE OF INVENTION: Immunostimulatory Compositions
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Ave., Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/146, 283
; FILING DATE: 03-SEPT-1998
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Judge, Linda R.

```

```

; REGISTRATION NUMBER: 42,702
; REFERENCE/DOCKET NUMBER: 7636-0010.21
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-324-0880
; TELEFAX: 650-324-0960
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 782 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: homo sapiens
; INDIVIDUAL ISOLATE: GM-CSF-Her-2 fusion protein; Fig. 8
US-09-146-283-4

Query Match 91.9%; Score 3632; DB 2; Length 782;
Best Local Similarity 98.7%; Pred. No. 1.9e-305;
Matches 659; Conservative 0; Mismatches 5; Indels 4; Gaps 2;

QY 1 METALCRMGILLALLPPGAASVQVCTGTDMLRLPASPEHDMRLHYQCGVQVGNL 60
METALCRMGILLALLPPGAASVQVCTGTDMLRLPASPEHDMRLHYQCGVQVGNL 60
DB 1 METALCRMGILLALLPPGAASVQVCTGTDMLRLPASPEHDMRLHYQCGVQVGNL 60
QY 61 ELTYLPNLSLFLQDIQEVGYVLIHNOVROVPLQRLRIYRGTOLEEDNVALAVDNG 120
ELTYLPNLSLFLQDIQEVGYVLIHNOVROVPLQRLRIYRGTOLEEDNVALAVDNG 120
DB 61 ELTYLPNLSLFLQDIQEVGYVLIHNOVROVPLQRLRIYRGTOLEEDNVALAVDNG 120
QY 121 DPLNNTPTVYGASPGGRLQRLSTLEIKGVLQIRNPOLCYODIILMDIFHKNNOLA 180
DPLNNTPTVYGASPGGRLQRLSTLEIKGVLQIRNPOLCYODIILMDIFHKNNOLA 180
DB 121 DPLNNTPTVYGASPGGRLQRLSTLEIKGVLQIRNPOLCYODIILMDIFHKNNOLA 180
QY 181 LTLIDTNRSRACHPCSPMCKSGRSGWESSSEDCSLRTVQAGCARGKGLPTDCHEOC 240
LTLIDTNRSRACHPCSPMCKSGRSGWESSSEDCSLRTVQAGCARGKGLPTDCHEOC 240
DB 181 LTLIDTNRSRACHPCSPMCKSGRSGWESSSEDCSLRTVQAGCARGKGLPTDCHEOC 240
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTEFSMNPBGRYTFGASCTYAC 300
AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTEFSMNPBGRYTFGASCTYAC 300
DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTEFSMNPBGRYTFGASCTYAC 300
QY 301 YNYLSTVGSCTLVCPHNOEVTAEEDTORCEKSKPCARVCGYGLMEHLREVRATYSAN 360
YNYLSTVGSCTLVCPHNOEVTAEEDTORCEKSKPCARVCGYGLMEHLREVRATYSAN 360
DB 301 YNYLSTVGSCTLVCPHNOEVTAEEDTORCEKSKPCARVCGYGLMEHLREVRATYSAN 360
QY 361 IOEFAGCKITFGSLAFIPESFDDPASNTAPLOPEQLQVEETLEETGYLYISAMPDSL 420
IOEFAGCKITFGSLAFIPESFDDPASNTAPLOPEQLQVEETLEETGYLYISAMPDSL 420
DB 361 IOEFAGCKITFGSLAFIPESFDDPASNTAPLOPEQLQVEETLEETGYLYISAMPDSL 420
QY 421 DLSVFQNLQVIRGRILHNGAYSILTGLGISTWLGSLRLSRLSGSLALIHNTHLCEVHTV 480
DLSVFQNLQVIRGRILHNGAYSILTGLGISTWLGSLRLSRLSGSLALIHNTHLCEVHTV 480
DB 421 DLSVFQNLQVIRGRILHNGAYSILTGLGISTWLGSLRLSRLSGSLALIHNTHLCEVHTV 480
QY 481 PMDQLEFRNHQALHTANRDEDECVGEGLACHOLCARHGCHMGSPFQVCVCSQFLRGQEC 540
PMDQLEFRNHQALHTANRDEDECVGEGLACHOLCARHGCHMGSPFQVCVCSQFLRGQEC 540
DB 481 PMDQLEFRNHQALHTANRDEDECVGEGLACHOLCARHGCHMGSPFQVCVCSQFLRGQEC 540
QY 541 VEECRVLOGLPREYVNAHRLCPHRECOPQNSVTCGPEADQCVACAHKDDPFVCAR 600
VEECRVLOGLPREYVNAHRLCPHRECOPQNSVTCGPEADQCVACAHKDDPFVCAR 600
DB 541 VEECRVLOGLPREYVNAHRLCPHRECOPQNSVTCGPEADQCVACAHKDDPFVCAR 600
QY 601 PSGVKPDLSTYMPIMKFPDEGACQPCPINCTHSCVDLDDKGCRAEDRASPLTISAVNG 660
PSGVKPDLSYMPIMKFPDEGACQPCPINCTHSCVDLDDKGCRAEDRASPLTISAVNG 660
DB 601 PSGVKPDLSTYMPIMKFPDEGACQPCPINCTHSCVDLDDKGCRAEDRASPLTISAVNG 660
QY 661 A-SPLDST 667
A-SPLDST 667
DB 661 A-SPLDST 665
A-SPLDST 665

RESULT 10
US-08-579-823A-4
; Sequence 4, Application US/08579823A

```

```
; Patent No. 6080409
; GENERAL INFORMATION:
; APPLICANT: Laus, Reiner
; APPLICANT: Ruegg, Curtis L.
; APPLICANT: Wu, Hongyu
; TITLE OF INVENTION: Immunostimulatory Composition and Method
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Ave. Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/579,823A
; FILING DATE: 03-DEC-1998
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Judge, Linda R.
; REGISTRATION NUMBER: 42,702
; REFERENCE/DOCKET NUMBER: 7636-0010
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-324-0960
; TELEFAX: 650-324-0960
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 782 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; .HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: homo sapiens
; INDIVIDUAL ISOLATE: GM-CSF-Her-2 fusion protein; Fig. 8
; US-08-579-823A-4

Query Match          91.9%; Score 3632; DB 3; Length 782;
Best Local Similarity 98.7%; Pred. No. 1.9e-305;
Matches 659; Conservative 0; Mismatches 5; Indels 4; Gaps 2;
```

```
QY 421 DLSVFQNLQVINGRIILHNGAYSLTLOGGLISWLGRLRELGSGALIHNNHLCFVHTV 480
| | | | |
DB 421 DLSVFQNLQVINGRIILHNGAYSLTLOGGLISWLGRLRELGSGALIHNNHLCFVHTV 480
QY 481 PWDOLFERNHQAHLTANRPEDECVEGELACHQLCARGHCWGPPTQCVNCSQFLRGDEC 540
| | | | |
DB 481 PWDOLFERNHQAHLTANRPEDECVEGELACHQLCARGHCWGPPTQCVNCSQFLRGDEC 540
QY 541 VEECVYLOGLPREYVNAHRCLEPCHECOPQNSVTCFGEADQCYACAHYKDPPECVARC 600
| | | | |
DB 541 VEECVYLOGLPREYVNAHRCLEPCHECOPQNSVTCFGEADQCYACAHYKDPPECVARC 600
QY 601 PSQVAPDLSTYMPIMKPEPEGACOPCPINCTHSCVDLDKGPAPORASPLTSQMEDDGP 660
| | | | |
DB 601 PSQVAPDLSTYMPIMKPEPEGACOPCPINCTHSCVDLDKGPAPORASPLTSLE---AP 657
QY 661 A-SPLDST 667
| | | | |
DB 658 ARSPSPST 665

RESULT 11
US-09-344-195-4
; Sequence 4, Application US/09344195
; Patent No. 6210662
; GENERAL INFORMATION:
; APPLICANT: Laus, Reiner
; APPLICANT: Ruegg, Curtis L.
; APPLICANT: Wu, Hongyu
; TITLE OF INVENTION: Immunostimulatory Compositions
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Ave. Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/344,195
; FILING DATE: 24-Jun-1999
; CLASSIFICATION: <unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/146,283
; FILING DATE: 03-SEPT-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Judge, Linda R.
; REGISTRATION NUMBER: 42,702
; REFERENCE/DOCKET NUMBER: 7636-0010.21
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-324-0960
; TELEFAX: 650-324-0960
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 782 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: homo sapiens
; INDIVIDUAL ISOLATE: GM-CSF-Her-2 fusion protein; Fig. 8
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:

US-09-344-195-4

Query Match          91.9%; Score 3632; DB 4; Length 782;
Best Local Similarity 98.7%; Pred. No. 1.9e-305;
```



Db 601 ACQPCPINCTHSCVLDLDDKGPAAE 624

RESULT 13

US-08-422-734-1

Sequence 1, Application US/08422734

Patent No. 6333169

GENERAL INFORMATION:

APPLICANT: Huddiak, Robert M.

APPLICANT: Shepard, H. Michael

APPLICANT: Ullrich, Axel

TITLE OF INVENTION: HER2 EXTRACELLULAR DOMAIN

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genentech, Inc.

STREET: 460 Point San Bruno Blvd

CITY: South San Francisco

STATE: California

COUNTRY: USA

ZIP: 94080

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: WinPatIn (Genentech)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/422,734

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/422108

FILING DATE: 14-Apr-1995

APPLICATION NUMBER: 08/355460

FILING DATE: 13-DEC-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/048346

FILING DATE: 15-APR-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/354319

FILING DATE: 19-MAY-1989

ATTORNEY/AGENT INFORMATION:

NAME: Lee, Wendy M

REGISTRATION NUMBER: 00,000

REFERENCE/DOCKET NUMBER: 554C2D1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415/225-1994

TELEFAX: 415/952-9881

TELEX: 910/371-7168

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 624 amino acids

TYPE: Amino Acid

TOPOLOGY: Linear

US-08-422-734-1

Query Match 87.8%; Score 3473; DB 4; Length 624;  
Best Local Similarity 99.8%; Pred. No. 7.7e-292;  
Matches 623; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 22 STQVCTGDMKRLPASETHLDMRLHLYOGCOVQGNLELYLPTNLSLFDIOEYO 81  
DB 1 STQVCTGDMKRLPASETHLDMRLHLYOGCOVQGNLELYLPTNLSLFDIOEYO 60  
QY 82 GYVLAHNOVROVPLQRLRIYVGTQLFEDNVALAVLDNGDPLNNTPTVTGASPGGIRELO 141  
DB 61 GYVLAHNOVROVPLQRLRIYVGTQLFEDNVALAVLDNGDPLNNTPTVTGASPGGIRELO 120  
QY 142 LRLSLLEIKGGVLIQNRNOLCYODITLTKDIFHKNNOLATLIDNRSRACHPCSPMKCG 201  
DB 121 LRLSLLEIKGGVLIQNRNOLCYODITLTKDIFHKNNOLATLIDNRSRACHPCSPMKCG 180  
QY 202 SRMGESSEDCSLTRTVACAGCARGKGPLPTDCHEOCAGCTGPKHSIDCLAHFNHS 261

Db 181 SRMGESSEDCSLTRTVACAGCARGKGPLPTDCHEOCAGCTGPKHSIDCLAHFNHS 240  
QY 262 GICEIHCAPALVYNTDFESMPNPEGRTYFGASCYVACPYNTLSTDVGSCTLVCPLANOE 321  
DB 241 GICEIHCAPALVYNTDFESMPNPEGRTYFGASCYVACPYNTLSTDVGSCTLVCPLANOE 300  
QY 322 VTAEDGTORCEKSPCARVCGIGLMEHLREVRATVSAIOEFACKKIFGSLAPLPSF 381  
DB 301 VTAEDGTORCEKSPCARVCGIGLMEHLREVRATVSAIOEFACKKIFGSLAPLPSF 360  
QY 382 DDDPASNTAPLOPEOLOVEFTELEITGYLYISAMPDLSPLDSVFNLOQYIRIRILHNGAY 441  
DB 361 DDDPASNTAPLOPEOLOVEFTELEITGYLYISAMPDLSPLDSVFNLOQYIRIRILHNGAY 420  
QY 442 SLTLOGISWGLRSLRLSGSLALIHNTHLCEVHTVPMQDLFRNPQALLHTANRPE 501  
DB 421 SLTLOGISWGLRSLRLSGSLALIHNTHLCEVHTVPMQDLFRNPQALLHTANRPE 480  
QY 502 DECVGEGLACHQLCARGHMGPGPTQCVNCSQFLNGQECVEECRYLGLPREYVNAHCL 561  
DB 481 DECVGEGLACHQLCARGHMGPGPTQCVNCSQFLNGQECVEECRYLGLPREYVNAHCL 540  
QY 562 PCHPCEQONGSVTCFGEADQVACAHYKDPFCVARGPSGVKPDLSYMPIMKPPDEEG 621  
DB 541 PCHPCEQONGSVTCFGEADQVACAHYKDPFCVARGPSGVKPDLSYMPIMKPPDEEG 600  
QY 622 ACQPCPINCTHSCVLDLDDKGPAAE 645  
DB 601 ACQPCPINCTHSCVLDLDDKGPAAE 624

RESULT 14  
US-09-630-155-2  
Sequence 2, Application US/09630155  
Patent No. 6414130

GENERAL INFORMATION:

APPLICANT: Doherty, Joni Kristin and Gail M. Clinton

TITLE OF INVENTION: HER-2 BINDING ANTAGONISTS

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESSEE: DAVIS WRIGHT TREMAINE LLP

STREET: 1501 Fourth Avenue, 2600 Century Square

CITY: Seattle

STATE: Washington

COUNTRY: U.S.A.

ZIP: 98101

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: PC compatible

OPERATING SYSTEM: Windows95

SOFTWARE: Word

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/630,155

FILING DATE: 16-Jan-2001

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Davison, Barry L.

REGISTRATION NUMBER: 47,309

REFERENCE/DOCKET NUMBER: 49321-10

TELECOMMUNICATION INFORMATION:

TELEPHONE: 206 628-7621

TELEFAX: 206 628-7699

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 419

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: unknown

MOLECULE TYPE: polypeptide

SEQUENCE DESCRIPTION: SEQ ID NO: 2:

US-09-630-155-2

Query Match 47.5%; Score 1878; DB 4; Length 419;  
Best Local Similarity 83.0%; Pred. No. 3.3e-154;  
Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;

QY 1 METALCRGGLLALLPFGAASVQCTGDMKRLPASPETHLMDRLHYGCGVQVGNL 60  
DB 1 METALCRGGLLALLPFGAASVQCTGDMKRLPASPETHLMDRLHYGCGVQVGNL 60  
QY 61 ELTYLPTNASLSPDIQIOEVGYVLLAHNQVROVPLQRLRIVRGTOLEFEDNVALVDNG 120  
DB 61 ELTYLPTNASLSPDIQIOEVGYVLLAHNQVROVPLQRLRIVRGTOLEFEDNVALVDNG 120  
QY 121 DPLNNTPTVTGASPGGIRELQRLSLTEILKGVLIQNPOLCYQDTILMDIFHKNNOLA 180  
DB 121 DPLNNTPTVTGASPGGIRELQRLSLTEILKGVLIQNPOLCYQDTILMDIFHKNNOLA 180  
QY 181 LTLIDNRSRACHPCSPKMGSGRCWGESSEDCSLTRTVAGGACRKGPLPDDCHEQC 240  
DB 181 LTLIDNRSRACHPCSPKMGSGRCWGESSEDCSLTRTVAGGACRKGPLPDDCHEQC 240  
QY 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDTEFSMPNREGRYTFGASCVTACP 300  
DB 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDTEFSMPNREGRYTFGASCVTACP 300  
QY 301 YNLTSDVGSCTLVCPDLHNOEVTAEQDQRCCKSPCARVYCYGLGMEHLREVAVTAN 360  
DB 301 YNLTSDVGSCTLVCPDLHNOEVTAEQDQRCCKSPCARVYCYGLGMEHLREVAVTAN 360  
QY 361 IOGFACCKRFGSLAPLEPESGDDPASNT---PPLQEQVQVETLEITGYLISAMPD 417  
DB 361 LKQPG--PAHPVLSFLRPSWDLVSAFYSLPLAPLSTSVPI-----SPVSVGRGPD 405  
QY 418 SLPLDSVFONLQVIRG 433  
DB 406 --PDAAVAVNLSTRYEG 419

RESULT 15  
US-08-484-438-7  
Sequence 7, Application US/08484438  
Patent No. 5811098  
Patent No. 5811098 5780031  
GENERAL INFORMATION:  
APPLICANT: Plowman, Gregory D.  
APPLICANT: Cuioussou, Jean-Michel  
APPLICANT: Shoyab, Mohammed  
APPLICANT: Siegall, Clay B.  
APPLICANT: Hellstet m, Ingegerd  
TITLE OF INVENTION: HER4 HUMAN RECEPTOR TYROSINE KINASE  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/484,438  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/323,442  
FILING DATE: 14-OCT-1994  
APPLICATION NUMBER: US 08/150,704  
FILING DATE: 10-NOV-1993  
CLASSIFICATION: 530

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/981,165  
FILING DATE: 24-NOV-1992  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 5624-230  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1210 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-484-438-7

Query Match 38.8%; Score 1534; DB 2; Length 1210;  
Best Local Similarity 42.9%; Pred. No. 7.7e-124;  
Matches 283; Conservative 109; Mismatches 240; Indels 28; Gaps 6;

QY 11 LLLALLPFGAA--STOYCTGDMKRLPASPETHLMDRLHYGCGVQVGNLPLETLPIN 68  
DB 14 LLAALCPASRALBEKKVCOQTSNKLTLQLEFDEHFLSLQRMFNCEVVLGNEITVQRN 73  
QY 69 ASLSFLQDIOEVGYVLLAHNQVROVPLQRLRIVRGTOLEFEDNVALVDNGPLNNTPT 128  
DB 74 YDLSFLKTOEVAGYVLLAHNQVROVPLQRLRIVRGTOLEFEDNVALVDNGPLNNTPT 126  
QY 129 VTGASPGGIRELQRLSLTEILKGVLIQNPOLCYQDTILMDIFHKNNOLA LTLIDN 188  
DB 127 ---ANKTGLKELPMRNQELHGAVRRSNNPALCNNEISQWRIVSSDLSNNSMQFNH 183  
QY 189 SRACHPCSPKMGSGRCWGESSEDCSLTRTVAGGACRKGPLPDDCHEQCAGCTGP 247  
DB 184 LGSCQKDPSCPNVSGWGAEEENCQKLTICAQCSGRCGRKSPSPDCCHNOCAGCTGP 243  
QY 248 KHSDCIACHFNHSGICELHCPALVYNTDTEFSMPNREGRYTFGASCVTACPYNLTSD 307  
DB 244 RESDCLVCRKFRDQATCKDPCPLMLNPTTYQMDVNPESKYSFGATCAKCRNTVYTD 303  
QY 308 VGSCTLVCPDLHNOEVTAEQDQRCCKSPCARVYCYGLGMEHLREVAVTANIOEPAGC 367  
DB 304 HGSCVRRACGADSYEM--BEDGVKCKCKECPGRKVCNIGIGEFKDSLSINATNIKHKNC 362  
QY 368 KTIFGSLAPLEPESGDDPASNTAPLOEQVETLEITGYLISAMPDLSLVFON 427  
DB 363 TSIISGDLHLPVAFRGDSFHTPLDQEDILKTYEITGFLIQAMPENRDLHAFEN 422  
QY 428 LQYIRGILNAGVSLTLOGISMLGLRSLRELGSLLALHNTHICLFVHTVPMQDLR 487  
DB 423 LEIRGRTHQGOFSIAVNSLNTSLGLRSLKLSDDVDVLTISGKNKLCVANTNMKFLG 482  
QY 488 NPQOALLHTANRDEDCVSGELACHOLCARGHCGMPPTQCVNCSQFLRGQECVEECRVL 547  
DB 483 TSGQKTIINRGENSKAIGQYCHALCSPEGCMGPRRCVSGRRNNSRRECVKDKKL 542  
QY 548 QGLPREVYNAHCLPCHPEQOPONGSYTCFPEADQCVACAHKDPFCVYARCPGKYPD 607  
DB 543 EGEPRFVENSECIOCHRECLPQAMNITTCGRPDNCIQCAHYIDGPHCVKTCGPAGVME 602  
QY 608 LSTYPIKFPDEBACQPCPINCCHSCVDDDKGCRAPQASPLTSONEDLGRASPDLST 667  
DB 603 NNTL-VMKYADAGHYCHLCHNCTYGTGGLRECCPTN-----GPKTBSIAT 648

Search completed: January 13, 2003, 14:50:01  
Job time: 22.2232 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 13, 2003, 14:47:00 : Search time 8.2943 Seconds  
(without alignments)  
1665.428 Million cell updates/sec

Title: US-09-854-356-7  
Perfect score: 3954  
Sequence: 1 MELAALCRWGLLLALLPFGA.....GFCPPDPAPGAGVHHRRH 712

Scoring table: BLOSUM62  
Gap 10.0 , Gapext 0.5

Searched: 118974 seqs, 19401057 residues

Total number of hits satisfying chosen parameters: 118974

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Published\_Applications\_AA:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEM\_PUB.pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCR\_NEM\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEM\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEM\_PUB.pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
8: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEM\_PUB.pep:\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEM\_PUB.pep:\*  
12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep:\*  
13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEM\_PUB.pep:\*  
14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3954	100.0	712	9	US-09-854-356-7
2	3954	100.0	919	9	US-09-854-356-6
3	3776	95.5	1255	9	US-09-854-356-1
4	3776	95.5	1255	9	US-09-930-125-2
5	3776	95.5	1255	10	US-09-811-123-9
6	3776	95.5	1255	10	US-09-811-115-3
7	3727	94.3	1255	9	US-09-769-508-2
8	3628	91.8	653	9	US-09-854-356-3
9	3590	90.8	645	10	US-09-921-161-1
10	3209.5	81.2	1256	9	US-09-854-356-2
11	3205	81.1	1260	9	US-09-870-759-118
12	3189.5	80.7	1256	9	US-09-854-356-14
13	3110.5	78.7	654	9	US-09-854-356-8
14	1614.5	40.8	564	10	US-09-821-883-3
15	1614.5	40.8	697	10	US-09-821-883-4
16	1612	40.8	555	10	US-09-821-883-1
17	1612	40.8	690	10	US-09-821-883-2
18	1608	40.7	479	10	US-09-821-883-5
19	1587	40.1	289	10	US-09-821-883-23

20	1534	38.8	1210	10	US-09-725-433-2	Sequence 2, Appl1
21	1451	36.7	1308	10	US-09-940-101-2	Sequence 2, Appl1
22	1447	36.6	615	10	US-09-940-101-4	Sequence 4, Appl1
23	1028.5	26.0	478	10	US-09-867-521-2	Sequence 2, Appl1
24	350	8.9	583	9	US-09-930-125-9	Sequence 9, Appl1
25	350	8.9	587	9	US-09-930-125-8	Sequence 8, Appl1
26	350	8.9	589	9	US-09-930-125-10	Sequence 10, Appl1
27	350	8.9	600	9	US-09-930-125-11	Sequence 11, Appl1
28	326	8.2	59	9	US-09-854-356-5	Sequence 5, Appl1
29	326	8.2	266	9	US-09-854-356-4	Sequence 4, Appl1
30	311.5	7.9	1367	9	US-09-870-759-120	Sequence 120, App
31	291	7.4	370	10	US-09-844-353A-104	Sequence 104, App
32	291	7.4	370	10	US-09-844-353A-104	Sequence 104, App
33	267.5	6.8	1724	10	US-09-844-353A-105	Sequence 105, App
34	267.5	6.8	1724	10	US-09-844-353A-106	Sequence 106, App
35	250	6.3	366	10	US-09-844-353A-103	Sequence 103, App
36	250	6.3	366	10	US-09-844-353A-105	Sequence 105, App
37	223.5	5.7	383	10	US-09-844-353A-105	Sequence 105, App
38	223.5	5.7	383	10	US-09-844-353A-105	Sequence 105, App
39	212	5.4	381	10	US-09-205-658-106	Sequence 106, App
40	212	5.4	381	10	US-09-844-353A-106	Sequence 106, App
41	161	4.1	3084	10	US-09-938-275-4	Sequence 4, Appl1
42	160	4.0	1111	10	US-09-756-071B-15	Sequence 15, Appl1
43	160	4.0	1193	10	US-09-756-071B-13	Sequence 13, Appl1
44	159	4.0	3635	10	US-09-845-583-2	Sequence 2, Appl1
45	154	3.9	830	9	US-09-870-759-134	Sequence 134, App

#### ALIGNMENTS

RESULT 1  
US-09-854-356-7  
: Sequence 7, Application US/09854356  
: Patent No. US2002017567A1  
: GENERAL INFORMATION:  
: APPLICANT: Cheever, Martin A.  
: APPLICANT: Coriysen, Dirk  
: APPLICANT: Smithkline Beecham Biologicals S. A.  
: TITLE OF INVENTION: HER-2/neu Fusion Proteins  
: FILE REFERENCE: 014058-009810PC  
: CURRENT APPLICATION NUMBER: US/09/854.356  
: CURRENT FILING DATE: 2001-05-09  
: PRIOR APPLICATION NUMBER: US 09/493.480  
: PRIOR FILING DATE: 2000-01-28  
: PRIOR APPLICATION NUMBER: US 60/117,976  
: PRIOR FILING DATE: 1999-01-29  
: NUMBER OF SEQ ID NOS: 26  
: SOFTWARE: PatentIn Ver. 2.1  
: SEQ ID NO 7  
: LENGTH: 712  
: TYPE: PRT  
: ORGANISM: Artificial Sequence  
: FEATURE:  
: OTHER INFORMATION: Description of Artificial Sequence: fusion protein  
: OTHER INFORMATION: of ECD and delta PD of human HER-2/neu  
US-09-854-356-7  
Query Match 100.0%: Score 3954; DB 9; Length 712;  
Best Local Similarity 100.0%: Pred. No. 1.6e-271;  
Matches 712; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MELAALCRWGLLLALLPFGAASVCTGTDKMLRIPASPERHLDMLRHLVYOCQVVGNTL 60  
DB 1 MELAALCRWGLLLALLPFGAASVCTGTDKMLRIPASPERHLDMLRHLVYOCQVVGNTL 60  
QY 61 ELTYPTNASSFLDDIDVGVYLAHNRQVPLQRLRYRGQLTFEDNATLAVLDNG 120  
DB 61 ELTYPTNASSFLDDIDVGVYLAHNRQVPLQRLRYRGQLTFEDNATLAVLDNG 120  
QY 121 DPLNNTVTGASPGLELDLRSITETIKGVLVLRNPOLCYODTIIMKDFHKNNOA 180  
DB 121 DPLNNTVTGASPGLELDLRSITETIKGVLVLRNPOLCYODTIIMKDFHKNNOA 180

Db	121	DPENNTTPTVTASRGSGRLERELDRSLTELKGVLIQGNRPQLCYDPTILMKDIPKNNOLA	180
Qy	181	LTLIDTNRSRACHPCSPCKSGRCWGESSEDCQSLRTTVACGAGCARCKGRLPTDCCHEQC	240
Db	181	LTLIDTNRSRACHPCSPCKSGRCWGESSEDCQSLRTTVACGAGCARCKGRLPTDCCHEQC	240
Qy	241	AAGCGPRHSQCLACIHNHNSGICTELHCPALVYNTPTFESMPRBERRTYFGASCYTACR	3000
Db	241	AAGCGPRHSQCLACIHNHNSGICTELHCPALVYNTPTFESMPRBERRTYFGASCYTACR	3000
Qy	301	YNYLSTDVSGCTIVCPRLNQCVTAEADGTQRCSEKCSKPCARCYGLGMEHLREYAVTYSAN	3606
Db	301	YNYLSTDVSGCTIVCPRLNQCVTAEADGTQRCSEKCSKPCARCYGLGMEHLREYAVTYSAN	3606
Qy	361	IQERAGCKKIGSLAFLEPESDGPASNTAPLQEPQLQVETTELTEITGYLYISAMPDSL	420
Db	361	IQERAGCKKIGSLAFLEPESDGPASNTAPLQEPQLQVETTELTEITGYLYISAMPDSL	420
Qy	421	DLVSQNQLQVIRGRILNHGAATSLTLOGIGISWMLRLSLRELGSGLAIHNHTHLCAPHTV	480
Db	421	DLVSQNQLQVIRGRILNHGAATSLTLOGIGISWMLRLSLRELGSGLAIHNHTHLCAPHTV	480
Qy	481	PMDQLFRNPQALLHTANRPDEDECVGEGSLACHQLCARGHCWGPRTQCVNCSQFLRGQEC	540
Db	481	PMDQLFRNPQALLHTANRPDEDECVGEGSLACHQLCARGHCWGPRTQCVNCSQFLRGQEC	540
Qy	541	VEECGVLOGGLREVYNANHCILCPREOPONGSVTCGPREADOCVAAAHNKDRPFCVARG	6000
Db	541	VEECGVLOGGLREVYNANHCILCPREOPONGSVTCGPREADOCVAAAHNKDRPFCVARG	6000
Qy	601	PSGVKPDLSYMPIMKFPDEGACQPCPRLNTHSCVDLDDKGCRAEORASPLTSQNEDLGP	6606
Db	601	PSGVKPDLSYMPIMKFPDEGACQPCPRLNTHSCVDLDDKGCRAEORASPLTSQNEDLGP	6606
Qy	661	ASPLDSTYRSLLEDDDKGLVDVAEEYLYPQGGFCDDPARGAGAMVHNHR	712
Db	661	ASPLDSTYRSLLEDDDKGLVDVAEEYLYPQGGFCDDPARGAGAMVHNHR	712
RESULT 2			
US-09-854-356-6			
Sequence 6, Application US/09854356			
Patent No. US20020177567A1			
GENERAL INFORMATION:			
APPLICANT: Cheever, Martin A.			
APPLICANT: Gheysen, Dirk			
APPLICANT: Corixa Corporation			
APPLICANT: SmithKline Beecham Biologicals S. A.			
TITLE OF INVENTION: HER-2/neu Fusion Proteins			
FILE REFERENCE: 014058-009810PC			
CURRENT APPLICATION NUMBER: US/09/854, 356			
CURRENT FILING DATE: 2001-05-09			
PRIOR APPLICATION NUMBER: US 09/493,480			
PRIOR FILING DATE: 2000-01-28			
PRIOR APPLICATION NUMBER: US 60/117, 976			
PRIOR FILING DATE: 1999-01-29			
NUMBER OF SEQ ID NOS: 26			
SOFTWARE: PatentIn Ver. 2.1			
SEQ ID NO 6			
LENGTH: 919			
TYPE: PRT			
ORGANISM: Artificial Sequence			
FEATURE:			
OTHER INFORMATION: Description of Artificial Sequence: fusion protein			
OTHER INFORMATION: of ECD and PD of human HER-2/neu			
US-09-854-356-6			

Query Match	100.0%	Score 3954;	DB 9;	Length 919;
Best Local Similarity	100.0%	Pred. No. 2.2e-271;		
Matches 712; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MELALCRGILLALLPGCAASTOYCTGTDMLRLPASETHLDMIRHLGYCCQVYQNL	60	

Dd	1	MEIALACSWGILLALILPPGASASTOVCTGDKMLRLPASPEHNDMLRHLYGCGVYVGNL	60
Oy	61	ELTYLPTNASLSFDLIDIOEVGYVLIAHNOVROPLORLRIVRGTOLEFDNYALAVLNG	120
Dd	61	ELTYLPTNASLSFDLIDIOEVGYVLIAHNOVROPLORLRIVRGTOLEFDNYALAVLNG	120
Oy	121	DPLNPTPVGTASRGGLRELOLRSLTELKGGVLIQNPOLCTYODTLIMKDIFKNNOLA	180
Dd	121	DPLNPTPVGTASRGGLRELOLRSLTELKGGVLIQNPOLCTYODTLIMKDIFKNNOLA	180
Oy	181	LTLIDTNRSRACHPCSPCKSRGCGWGESSEEGCSLFTFTVCAGGCAKCGPLPTDCHEOC	240
Dd	181	LTLIDTNRSRACHPCSPCKSRGCGWGESSEEGCSLFTFTVCAGGCAKCGPLPTDCHEOC	240
Oy	241	AAGCTGPKHSDCLACLFHNHSGICELHCPALVYNTDTFESMPNEBGRYTFGASCVTACP	300
Dd	241	AAGCTGPKHSDCLACLFHNHSGICELHCPALVYNTDTFESMPNEBGRYTFGASCVTACP	300
Oy	301	YNYLSTDVSGCTIYCPHLHNOEVTAEADGTORCEKSKPCARCYGLGEMHLREVAAYTSAN	360
Dd	301	YNYLSTDVSGCTIYCPHLHNOEVTAEADGTORCEKSKPCARCYGLGEMHLREVAAYTSAN	360
Oy	361	IOEFAGCKKIGSLAFLPESFSDGDPASNTAPLQEOLOVEFTEIEITGYLIYSAMPDSL	420
Dd	361	IOEFAGCKKIGSLAFLPESFSDGDPASNTAPLQEOLOVEFTEIEITGYLIYSAMPDSL	420
Oy	421	DLVSFQNLQVLRGRILNHGAATLTLQGLGSLWLSRLSRELSSGLALIHNTNHCLEFHTV	480
Dd	421	DLVSFQNLQVLRGRILNHGAATLTLQGLGSLWLSRLSRELSSGLALIHNTNHCLEFHTV	480
Oy	481	PMDOLFRRPQHALLHTANRPDEECVGEGLACHOLCARGHCGPRTQCVNCSOFLRGQEC	540
Dd	481	PMDOLFRRPQHALLHTANRPDEECVGEGLACHOLCARGHCGPRTQCVNCSOFLRGQEC	540
Oy	541	VEECVLOGLEPREVYNNAHCLCPHRECOQPOGSGYTCGPREADQCVAAHNTKDRPFCVARC	600
Dd	541	VEECVLOGLEPREVYNNAHCLCPHRECOQPOGSGYTCGPREADQCVAAHNTKDRPFCVARC	600
Oy	601	PSGVPRDLSTYMPIMKFPDEBACQPCPINTCHSCVDLDDKGCAPAEORASPLTSQNEIDLGP	660
Dd	601	PSGVPRDLSTYMPIMKFPDEBACQPCPINTCHSCVDLDDKGCAPAEORASPLTSQNEIDLGP	660
Oy	661	ASPLDSTYRSRLBEDDDMGDLVDAEBEYLVPOOGFFCCDPAPAGAGMYHNHRH	712
Dd	661	ASPLDSTYRSRLBEDDDMGDLVDAEBEYLVPOOGFFCCDPAPAGAGMYHNHRH	712
RESULT 3			
US-09-854-356-1			
: Sequence 1, Application US/09854356			
: Patent No. US20020177567A1			
: GENERAL INFORMATION:			
: APPLICANT: Cheever, Martin A.			
: APPLICANT: Gheysen, Dirk			
: APPLICANT: Corixa Corporation			
: APPLICANT: SmithKline Beecham Biologicals S. A.			
: TITLE OF INVENTION: HER-2/neu Fusion Proteins			
: FILE REFERENCE: 014058-009810PC			
: CURRENT APPLICATION NUMBER: US/09/854, 356			
: CURRENT FILING DATE: 2001-05-09			
: PRIOR APPLICATION NUMBER: US 09/493,480			
: PRIOR FILING DATE: 2000-01-28			
: PRIOR APPLICATION NUMBER: US 60/117,976			
: PRIOR FILING DATE: 1999-01-29			
: NUMBER OF SEQ ID NOS: 26			
: SOFTWARE: PatentIn Ver. 2.1			

RESULT 3  
 : US-09-854-356-1  
 : Sequence 1, Application US/09854356  
 : Patent No. US2002017567A1  
 : GENERAL INFORMATION:  
 : APPLICANT: Cheever, Martin A.  
 : APPLICANT: Gheysen, Dirk  
 : APPLICANT: Corixa Corporation  
 : APPLICANT: SmithKline Beecham Biologicals  
 : TITLE OF INVENTION: HER-2/neu Fusion Protein  
 : FILE REFERENCE: 014058-009810PC  
 : CURRENT APPLICATION NUMBER: US/09/854,356  
 : CURRENT FILING DATE: 2001-05-09  
 : PRIOR APPLICATION NUMBER: US 09/493,480  
 : PRIOR FILING DATE: 2000-01-28  
 : PRIOR APPLICATION NUMBER: US 60/117,976  
 : PRIOR FILING DATE: 1999-01-29  
 : NUMBER OF SEQ ID NOS: 26  
 : SOFTWARE: PatentIn Ver. 2.1  
 : SEQ ID NO 1  
 : LENGTH: 1255  
 : TYPE: prt  
 : ORGANISM: Homo sapiens  
 : FEATURE:  
 : OTHER INFORMATION: human HER-2/neu protein  
 : NAME/KEY: DOMAIN



LOCATION: (1)..(653)  
OTHER INFORMATION: extracellular domain (ECD)  
NAME/KEY: DOMAIN  
LOCATION: (676)..(1255)  
OTHER INFORMATION: Intracellular domain (ICD)  
NAME/KEY: DOMAIN  
LOCATION: (990)..(1255)  
OTHER INFORMATION: phosphorylation domain (PD)  
NAME/KEY: DOMAIN  
LOCATION: (990)..(1048)  
OTHER INFORMATION: fragment of the phosphorylation domain, preferred  
OTHER INFORMATION: portion (delta PD)  
US-09-854-356-1

Query Match 95.5%; Score 3776; DB 9; Length 1255;  
Best Local Similarity 67.9%; Pred. No. 1.2e-258;  
Matches 712; Conservative 0; Mismatches 0; Indels 336; Gaps 1;

QY 1 MELALCRGGLLALLPPAASVCTGDMKRLPASPETHLMDRLHYOGCQVYQGNL 60  
DB 1 MELALCRGGLLALLPPAASVCTGDMKRLPASPETHLMDRLHYOGCQVYQGNL 60  
QY 61 ELYLPTNASLSFLDIOEVQGYVLIHNOVQVPLQRLIRIVGTQLFEDNTALAVLDNG 120  
DB 61 ELYLPTNASLSFLDIOEVQGYVLIHNOVQVPLQRLIRIVGTQLFEDNTALAVLDNG 120  
QY 121 DPLNNTPTVTGASPGGLRLQRLSLTEILKGVLIQVLPOLCYODTILMKDIFHKNNOLA 180  
DB 121 DPLNNTPTVTGASPGGLRLQRLSLTEILKGVLIQVLPOLCYODTILMKDIFHKNNOLA 180  
QY 181 LTLIDNRSRACHPCSPMKSGSRGSESESDOSLTRVCAGACARCKPPLTDCCHQC 240  
DB 181 LTLIDNRSRACHPCSPMKSGSRGSESESDOSLTRVCAGACARCKPPLTDCCHQC 240  
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMPNPEGRTYFGASCVTACP 300  
DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMPNPEGRTYFGASCVTACP 300  
QY 301 YNYLSTDVGSCTLVCPRLHNOEYTAEDGTQRCCKSPCARVCYGLGMEHLREVRAVTSAN 360  
DB 301 YNYLSTDVGSCTLVCPRLHNOEYTAEDGTQRCCKSPCARVCYGLGMEHLREVRAVTSAN 360  
QY 361 IOEPAGCKKIFGSLAFLPSPFDGDPASNTAPLOPQLOVFEETLEETGTYLSAMPDSL 420  
DB 361 IOEPAGCKKIFGSLAFLPSPFDGDPASNTAPLOPQLOVFEETLEETGTYLSAMPDSL 420  
QY 421 DLSVFQNLQVIRGRILHNGAVSLTLOGLGISWLGRLSRLSGSLALIHNTHLCEVHTV 480  
DB 421 DLSVFQNLQVIRGRILHNGAVSLTLOGLGISWLGRLSRLSGSLALIHNTHLCEVHTV 480  
QY 481 PWDQLFRNPHOALLHTANRDECEVGEGLAQHQLCARHGWGPRPOCVNCSOFLRGQRC 540  
DB 481 PWDQLFRNPHOALLHTANRDECEVGEGLAQHQLCARHGWGPRPOCVNCSOFLRGQRC 540  
QY 541 VECEKVLQGLPREYVNAHRCPCHPCCOPONGSVTCFGEAOCVACAYKDPPECVAC 600  
DB 541 VECEKVLQGLPREYVNAHRCPCHPCCOPONGSVTCFGEAOCVACAYKDPPECVAC 600  
QY 601 PSGVPRDLSYMPIMKFPDEGACOPCPICTHSCVDLDKGPABEGRASPLSIISAVYG 660  
DB 601 PSGVPRDLSYMPIMKFPDEGACOPCPICTHSCVDLDKGPABEGRASPLSIISAVYG 660  
QY 654 ----- 653  
DB 661 ILVVVGLVFGILIRKROQKIRKTYMRRLQETELVEPLTPSGAMPNOAQRIKLETEL 720  
QY 654 ----- 653  
DB 721 RKVKVLGSGAFGTIVKGIWIPGENVKIPVAILKVLRENTSPANKETLDEAYVAGVGP 780  
QY 654 ----- 653  
DB 781 YVSRLLGICLTSTVOLVTPQMLPYGCLLDHVRNRRGLSQDILLNMCQIAKMSYLEVDR 840

QY 654 ----- 653  
DB 841 LVHRDLAARNLVKSPBNHVKITDFGLARLLIDETEYHADGKVPILKMALESILRRFT 900  
QY 654 ----- 653  
DB 901 HOSDVMSYGVWELMTFPAKPYDGIIPAREIDPLEKGERLPQPICTIDYIMIMYKCM 960  
QY 654 ----- 1021  
DB 961 IDSECRPRELVESEFSRMAPDPQRFVQIONEDLGPASPLDSTFRSLLEDMDGLVDA 1020  
QY 685 EEYVPOGFFCQDPDAPAGAGVHHNR 712  
DB 1021 EEYVPOGFFCQDPDAPAGAGVHHNR 1048

RESULT 4  
US-09-930-125-2  
Sequence 2, Application US/09930125  
Publication No. US20020193329A1  
GENERAL INFORMATION:  
APPLICANT: Hand-Zimmerman, Susan  
APPLICANT: Cheever, Martin A.  
APPLICANT: Foy, Teresa M.  
APPLICANT: Lodges, Michael J.  
APPLICANT: Kalos, Michael D.  
APPLICANT: McNeill, Patricia D.  
APPLICANT: Vedvick, Thomas S.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND DIAGNOSIS  
OF HER-2/NEU-ASSOCIATED MALIGNANCIES  
FILE REFERENCE: 210121.544  
CURRENT APPLICATION NUMBER: US/09/930,125  
CURRENT FILING DATE: 2001-08-14  
NUMBER OF SEQ ID NOS: 25  
SOFTWARE: FastSeq for Windows version 3.0  
SEQ ID NO 2  
LENGTH: 1255  
TYPE: PRT  
ORGANISM: Homo sapien  
US-09-930-125-2

Query Match 95.5%; Score 3776; DB 9; Length 1255;  
Best Local Similarity 67.9%; Pred. No. 1.2e-258;  
Matches 712; Conservative 0; Mismatches 0; Indels 336; Gaps 1;

QY 1 MELALCRGGLLALLPPAASVCTGDMKRLPASPETHLMDRLHYOGCQVYQGNL 60  
DB 1 MELALCRGGLLALLPPAASVCTGDMKRLPASPETHLMDRLHYOGCQVYQGNL 60  
QY 61 ELYLPTNASLSFLDIOEVQGYVLIHNOVQVPLQRLIRIVGTQLFEDNTALAVLDNG 120  
DB 61 ELYLPTNASLSFLDIOEVQGYVLIHNOVQVPLQRLIRIVGTQLFEDNTALAVLDNG 120  
QY 121 DPLNNTPTVTGASPGGLRLQRLSLTEILKGVLIQVLPOLCYODTILMKDIFHKNNOLA 180  
DB 121 DPLNNTPTVTGASPGGLRLQRLSLTEILKGVLIQVLPOLCYODTILMKDIFHKNNOLA 180  
QY 181 LTLIDNRSRACHPCSPMKSGSRGSESESDOSLTRVCAGACARCKPPLTDCCHQC 240  
DB 181 LTLIDNRSRACHPCSPMKSGSRGSESESDOSLTRVCAGACARCKPPLTDCCHQC 240  
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMPNPEGRTYFGASCVTACP 300  
DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMPNPEGRTYFGASCVTACP 300  
QY 301 YNYLSTDVGSCTLVCPRLHNOEYTAEDGTQRCCKSPCARVCYGLGMEHLREVRAVTSAN 360  
DB 301 YNYLSTDVGSCTLVCPRLHNOEYTAEDGTQRCCKSPCARVCYGLGMEHLREVRAVTSAN 360  
QY 361 IOEPAGCKKIFGSLAFLPSPFDGDPASNTAPLOPQLOVFEETLEETGTYLSAMPDSL 420  
DB 361 IOEPAGCKKIFGSLAFLPSPFDGDPASNTAPLOPQLOVFEETLEETGTYLSAMPDSL 420

Db 361 IOEFAGCKITFSLAFIPESFDGDPASNTAPLOPEOLOVFEETLEITGYLISAMPDLP 420  
QY 421 DLVSFQMLQVIRGRILNNGAVSLTLOGISWGLRSRLREGSGALIHNTHTLCFVHTV 480  
Db 421 DLVSFQMLQVIRGRILNNGAVSLTLOGISWGLRSRLREGSGALIHNTHTLCFVHTV 480  
QY 481 PMDLFRNPQALHTANRPEDCEVSGELACHQCLCARGHCHGMPPTOCVNCISOFLRGQEC 540  
Db 481 PMDLFRNPQALHTANRPEDCEVSGELACHQCLCARGHCHGMPPTOCVNCISOFLRGQEC 540  
QY 541 VEECRVLOGLPREYVNAHCLPCHECOPONGSYTCGPEADQCVACAHYKDPFCVARC 600  
Db 541 VEECRVLOGLPREYVNAHCLPCHECOPONGSYTCGPEADQCVACAHYKDPFCVARC 600  
QY 601 PSQVKPDLSTYMPWKPFDEBEGACQPCPINCTHSCVDLDDKCPAEOBASPPTS----- 653  
Db 601 PSQVKPDLSTYMPWKPFDEBEGACQPCPINCTHSCVDLDDKCPAEOBASPPTS----- 653  
QY 654 ----- 653  
Db 661 ILVVVLGVVFGILIKRROQKIRKYMRLLOETELVEPLTPSGAMPNOAMRIKETEEL 720  
QY 654 ----- 653  
Db 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSFKANKELIDEAYVNAVGSF 780  
QY 654 ----- 653  
Db 781 YVSRLLGICLSTVQVLTQMLPFGCLLDHVRENKRLGSDLLMWCMQIAKMSYLEDR 840  
QY 654 ----- 653  
Db 841 LVHRDLAARNLVKSPNHVKITDFGLARLDDIDETEHADGKVPKMALESILRRFT 900  
QY 654 ----- 653  
Db 901 HOSDVWSTGYTVWELMTFFGAKPYDGIAPARETIDLEKGERLPORPCTIDVYMTMVKCM 960  
QY 654 ----- -QNEDLGPASPLDSTFYRSLEDDMDGLVDA 684  
Db 961 IDSECRPRFRELYSEFSMARDPQRFVYIQNEDLGPAASPDLSTFYRSLEDDMDGLVDA 1020  
QY 685 EEYLVPOOGFFCPDPAPGAGMHHRR 712  
Db 1021 EEYLVPOOGFFCPDPAPGAGMHHRR 1048

RESULT 5  
US-09-811-123-9  
; Sequence 9, Application US/09811123  
; Patent No. US2002001587A1  
; GENERAL INFORMATION:  
; APPLICANT: Sharon Erickson  
; APPLICANT: Ralph Schwall  
; APPLICANT: Mark Sliwowski  
; TITLE OF INVENTION: METHODS OF TREATMENT USING ANTI-ETDB  
; FILE REFERENCE: GENENT. 073A2  
; CURRENT APPLICATION NUMBER: US/09/811,123  
; CURRENT FILING DATE: 2001-03-16  
; PRIOR APPLICATION NUMBER: 60/238,327  
; PRIOR FILING DATE: 2000-10-05  
; PRIOR APPLICATION NUMBER: 09/602,530  
; PRIOR FILING DATE: 2000-06-23  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 9  
; LENGTH: 1255  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-811-123-9

Query Match 95.5%; Score 3776; DB 10; Length 1255;

Best Local Similarity 67.9%; Pred. No. 1,2e-258;  
Matches 712; Conservative 0; Mismatches 0; Indels 336; Gaps 1;  
QY 1 MELALACRMGILLALLPPGAASVCTGTGDKKLPLASPEHNDMLKHLVQGVQVGNL 60  
Db 1 MELALACRMGILLALLPPGAASVCTGTGDKKLPLASPEHNDMLKHLVQGVQVGNL 60  
QY 61 ELTYLPTNASLSFLODIOEVQGYVLIAHNGVQVPLRLRIVRSTQOLFEDNYALAVDNG 120  
Db 61 ELTYLPTNASLSFLODIOEVQGYVLIAHNGVQVPLRLRIVRSTQOLFEDNYALAVDNG 120  
QY 121 DPLNNTPTVTCASPGGLREIQLRSLTEILKGVLIQBNPOLCYODTILMKDIFKNNOLA 180  
Db 121 DPLNNTPTVTCASPGGLREIQLRSLTEILKGVLIQBNPOLCYODTILMKDIFKNNOLA 180  
QY 181 LTLIDTNRSRACHPCSPCKGSRMGSSSEDCOSTLTFTVCAGGACRKGPLPTCCHEOC 240  
Db 181 LTLIDTNRSRACHPCSPCKGSRMGSSSEDCOSTLTFTVCAGGACRKGPLPTCCHEOC 240  
QY 241 AAGCTGPKHSDCLACLFHNHSGICELCPALVTYNTDTFESMPNDEGRYTFGASCYACP 300  
Db 241 AAGCTGPKHSDCLACLFHNHSGICELCPALVTYNTDTFESMPNDEGRYTFGASCYACP 300  
QY 301 YNYLSTDVGSCTLVCPILANDVTAEDGTQRCCKSKPCARVCTGLGHEHLEREVAATVSAN 360  
Db 301 YNYLSTDVGSCTLVCPILANDVTAEDGTQRCCKSKPCARVCTGLGHEHLEREVAATVSAN 360  
QY 361 IOEFAGCKITFSLAFIPESFDGDPASNTAPLOPEOLOVFEETLEITGYLISAMPDLP 420  
Db 361 IOEFAGCKITFSLAFIPESFDGDPASNTAPLOPEOLOVFEETLEITGYLISAMPDLP 420  
QY 421 DLVSFQMLQVIRGRILNNGAVSLTLOGISWGLRSRLREGSGALIHNTHTLCFVHTV 480  
Db 421 DLVSFQMLQVIRGRILNNGAVSLTLOGISWGLRSRLREGSGALIHNTHTLCFVHTV 480  
QY 481 PMDLFRNPQALHTANRPEDCEVSGELACHQCLCARGHCHGMPPTOCVNCISOFLRGQEC 540  
Db 481 PMDLFRNPQALHTANRPEDCEVSGELACHQCLCARGHCHGMPPTOCVNCISOFLRGQEC 540  
QY 541 VEECRVLOGLPREYVNAHCLPCHECOPONGSYTCGPEADQCVACAHYKDPFCVARC 600  
Db 541 VEECRVLOGLPREYVNAHCLPCHECOPONGSYTCGPEADQCVACAHYKDPFCVARC 600  
QY 601 PSQVKPDLSTYMPWKPFDEBEGACQPCPINCTHSCVDLDDKCPAEOBASPPTS----- 653  
Db 601 PSQVKPDLSTYMPWKPFDEBEGACQPCPINCTHSCVDLDDKCPAEOBASPPTS----- 653  
QY 654 ----- 653  
Db 661 ILVVVLGVVFGILIKRROQKIRKYMRLLOETELVEPLTPSGAMPNOAMRIKETEEL 720  
QY 654 ----- 653  
Db 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSFKANKELIDEAYVNAVGSF 780  
QY 654 ----- 653  
Db 781 YVSRLLGICLSTVQVLTQMLPFGCLLDHVRENKRLGSDLLMWCMQIAKMSYLEDR 840  
QY 654 ----- 653  
Db 841 LVHRDLAARNLVKSPNHVKITDFGLARLDDIDETEHADGKVPKMALESILRRFT 900  
QY 654 ----- 653  
Db 901 HOSDVWSTGYTVWELMTFFGAKPYDGIAPARETIDLEKGERLPORPCTIDVYMTMVKCM 960  
QY 654 ----- -QNEDLGPASPLDSTFYRSLEDDMDGLVDA 684  
Db 961 IDSECRPRFRELYSEFSMARDPQRFVYIQNEDLGPAASPDLSTFYRSLEDDMDGLVDA 1020  
QY 685 EEYLVPOOGFFCPDPAPGAGMHHRR 712

DB 1021 EEYLVPOGFCPPDPAPGAGMVHHRH 1048

RESULT 6  
US-09-811-115-3  
; Sequence 3, Application US/09811115  
; Patent No. US20020035736A1  
; GENERAL INFORMATION:  
; APPLICANT: Erickson, Sharon  
; APPLICANT: Schwall, Ralph  
; APPLICANT: King, Kathleen  
; TITLE OF INVENTION: HER-2 TRANSGENIC NON-HUMAN TUMOR MODEL  
; FILE REFERENCE: GENE 034A  
; CURRENT APPLICATION NUMBER: US/09/811,115  
; PRIOR FILING DATE: 2001-03-16  
; PRIOR APPLICATION NUMBER: 60/189,844  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FASTSEQ for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 1255  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-811-115-3

Query Match 95.5%; Score 3776; DB 10; Length 1255;  
Best Local Similarity 67.9%; Pred. No. 1.2e-258;  
Matches 712; Conservative 0; Mismatches 0; Indels 336; Gaps 1;

QY 1 MELALCRWGLLALLPPGAASVQCTGTDMLRLPASPETHLDMRLHLYOGCQVQGNL 60  
DB 1 MELALCRWGLLALLPPGAASVQCTGTDMLRLPASPETHLDMRLHLYOGCQVQGNL 60  
QY 61 ETTYPTNASLSFLODIOEVQGYVLIANOVQVPLQRLRIYRGQLEFEDNYALAVLNG 120  
DB 61 ETTYPTNASLSFLODIOEVQGYVLIANOVQVPLQRLRIYRGQLEFEDNYALAVLNG 120  
QY 121 DPLNNTFVTGASPGGLRELOLRSLTEILKGVLIQNRNPQLCYOPTILMKDIFHKNOLA 180  
DB 121 DPLNNTFVTGASPGGLRELOLRSLTEILKGVLIQNRNPQLCYOPTILMKDIFHKNOLA 180  
QY 181 LTLIDTNSRACHPCSPCKSGRCWGESSEDCOSLTRTYCAGGACRCKGRLPTDCHEOC 240  
DB 181 LTLIDTNSRACHPCSPCKSGRCWGESSEDCOSLTRTYCAGGACRCKGRLPTDCHEOC 240  
QY 241 AAGCTGPRHSDCLACIHFHNSGICELHCPALVTYNTDTFESMPNEGRYTFGASCVTACP 300  
DB 241 AAGCTGPRHSDCLACIHFHNSGICELHCPALVTYNTDTFESMPNEGRYTFGASCVTACP 300  
QY 301 YNYLSTDVGSCTVLCPLNNOEYTAEDGTQRCCKSPCARVCYGLGMEHLRVRVAVTSAN 360  
DB 301 YNYLSTDVGSCTVLCPLNNOEYTAEDGTQRCCKSPCARVCYGLGMEHLRVRVAVTSAN 360  
QY 361 IOEFAGCKKIFGSLAFLEPSEFGDPAWNTAPLQPEOLQVFEETLEITGVLYSANPDSL 420  
DB 361 IOEFAGCKKIFGSLAFLEPSEFGDPAWNTAPLQPEOLQVFEETLEITGVLYSANPDSL 420  
QY 421 DLSVFQNTQVIRGRIHNGAVSLTQGLGISMGLRSLRELGSGIALIHNHTLCEFTV 480  
DB 421 DLSVFQNTQVIRGRIHNGAVSLTQGLGISMGLRSLRELGSGIALIHNHTLCEFTV 480  
QY 481 PMDOLFRRPHALLHTARPEDECEGELACHOICARHCWCPRTQCVNCSQFLRGEC 540  
DB 481 PMDOLFRRPHALLHTARPEDECEGELACHOICARHCWCPRTQCVNCSQFLRGEC 540  
QY 541 VEECRVLOGLPREVYNAHCLPCHEPCOPONGSVTCFEPREADQCYACAHYKDPFCVARC 600  
DB 541 VEECRVLOGLPREVYNAHCLPCHEPCOPONGSVTCFEPREADQCYACAHYKDPFCVARC 600  
QY 601 PSQVAPDLSTYPIKFPDEBEGACOPCLNCTHSCVDLDDKGPABQASPLTS----- 653  
DB 601 PSQVAPDLSTYPIKFPDEBEGACOPCLNCTHSCVDLDDKGPABQASPLTS----- 653

QY 654 ----- 653  
DB 661 ILLVYVAGVVEGILLKRRQKIRKTYMRLLOETELVERLTPSGAMPNQAQRILKETEL 720  
QY 654 ----- 653  
DB 721 RKVVLGSGAFGTYYKGIWIPDGENVKIPAIKVLRENTSPKANKEILDEAVYVAGVSP 780  
QY 654 ----- 653  
DB 781 YVSRLLGICLSTVOLVQLMPYGLLDHVRNRRGLSGODLNMCMQIAKMSYLEDVR 840  
QY 654 ----- 653  
DB 841 LVHRLAARNVLKSPNHHKITDPLGLARLLIDETEHADGKVPDKMALESLIRRET 900  
QY 654 ----- 653  
DB 901 HQSDVSYGVTVWELMTFGAKPYDIPAREIPDLLEKGERLPQPICTIDVYMIWKCMM 960  
QY 654 ----- 653  
DB 961 IDSECRPRFRELVSFSESRMARDPQRFVYIQNEDLGPASPLDSTFYRSLLEDMDGLVDA 1020  
QY 685 EEYLVPOGFCPPDPAPGAGMVHHRH 712  
DB 1021 EEYLVPOGFCPPDPAPGAGMVHHRH 1048

RESULT 7  
US-09-769-508-2  
; Sequence 2, Application US/09769508  
; Patent No. US20020155527A1  
; GENERAL INFORMATION:  
; APPLICANT: STUART, SUSAN G.  
; APPLICANT: MONAHAN, JOHN J.  
; APPLICANT: LANGTON, BEATRICE CLAUDIA  
; APPLICANT: HANCOCK, MIRIAM E.C.  
; APPLICANT: CHAO, LORRINE A.  
; TITLE OF INVENTION: C-ERBB-2 EXTERNAL DOMAIN: GP75  
; FILE REFERENCE: BEBIO-111-C1  
; CURRENT APPLICATION NUMBER: US/09/769,508  
; CURRENT FILING DATE: 2001-01-26  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 1255  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-769-508-2

Query Match 94.3%; Score 3727; DB 9; Length 1255;  
Best Local Similarity 67.5%; Pred. No. 3.3e-255;  
Matches 707; Conservative 0; Mismatches 5; Indels 336; Gaps 1;

QY 1 MELALCRWGLLALLPPGAASVQCTGTDMLRLPASPETHLDMRLHLYOGCQVQGNL 60  
DB 1 MELALCRWGLLALLPPGAASVQCTGTDMLRLPASPETHLDMRLHLYOGCQVQGNL 60  
QY 61 ETTYPTNASLSFLODIOEVQGYVLIANOVQVPLQRLRIYRGQLEFEDNYALAVLNG 120  
DB 61 ETTYPTNASLSFLODIOEVQGYVLIANOVQVPLQRLRIYRGQLEFEDNYALAVLNG 120  
QY 121 DPLNNTFVTGASPGGLRELOLRSLTEILKGVLIQNRNPQLCYOPTILMKDIFHKNOLA 180  
DB 121 DPLNNTFVTGASPGGLRELOLRSLTEILKGVLIQNRNPQLCYOPTILMKDIFHKNOLA 180  
QY 181 LTLIDTNSRACHPCSPCKSGRCWGESSEDCOSLTRTYCAGGACRCKGRLPTDCHEOC 240  
DB 181 LTLIDTNSRACHPCSPCKSGRCWGESSEDCOSLTRTYCAGGACRCKGRLPTDCHEOC 240  
QY 241 AAGCTGPRHSDCLACIHFHNSGICELHCPALVTYNTDTFESMPNEGRYTFGASCVTACP 300

```

Db 241 AAGCTGKHSDDCLACLFHNSGICELCPALVYNTDTFESMPNPEGRYTFGASCYTACP 300
QY 301 YNTLSTVGSCTLVCPPLHNOEVAEDGTORCEKSKPCARVCYGLGMEHLREVAAYTSAN 360
Db 301 YNTLSTVGSCTLVCPPLHNOEVAEDGTORCEKSKPCARVCYGLGMEHLREVAAYTSAN 360
QY 361 IOEFAGCKITFGSLAFIPESFDGDPASNTAPLOPEOLQVEETELETGYLYISAMPDLP 420
Db 361 IOEFAGCKITFGSLAFIPESFDGDPASNTAPLOPEOLQVEETELETGYLYISAMPDLP 420
QY 421 DLSVFQMLQYIRGRILHNGAYSLTLQGLISWLGRLSRLGSLALIHNTHLCEVHTV 480
Db 421 DLSVFQMLQYIRGRILHNGAYSLTLQGLISWLGRLSRLGSLALIHNTHLCEVHTV 480
QY 481 PMDOLFNPQALHTANRREDECVGEGSLACHOLCARGHCGMPGPTQCVNCSOFLRGQEC 540
Db 481 PMDOLFNPQALHTANRREDECVGEGSLACHOLCARGHCGMPGPTQCVNCSOFLRGQEC 540
QY 541 VEECRVLQGLPREVYNARHCLPCHPECOPONGSVTCGFPADQCVACAHYKDPFCVARC 600
Db 541 VEECRVLQGLPREVYNARHCLPCHPECOPONGSVTCGFPADQCVACAHYKDPFCVARC 600
QY 601 PSQVKPDLSTYMPIMKFPDEEGACQPCPINCTHSCVDLDKGCRAEQASPLTS 653
Db 601 PSQVKPDLSTYMPIMKFPDEEGACQPCPINCTHSCVDLDKGCRAEQASPLTS 653
QY 654 ILVVVGVVFGILIRKROOKIRKTYMRLLQETELVEPLTPSGAMPNOAMRIKETEL 720
Db 654 ILVVVGVVFGILIRKROOKIRKTYMRLLQETELVEPLTPSGAMPNOAMRIKETEL 720
QY 721 RKVKVLGSGAFGVYKGIWIPDGENYKIPVAIKVLRNTSPKANKETLDEAYVAVGSP 780
Db 721 RKVKVLGSGAFGVYKGIWIPDGENYKIPVAIKVLRNTSPKANKETLDEAYVAVGSP 780
QY 781 YSRLLGICLTSTVQVLTQMPYGCLLDHVRENRLGSLDLMCMQIAKMSYLEDDR 840
Db 781 YSRLLGICLTSTVQVLTQMPYGCLLDHVRENRLGSLDLMCMQIAKMSYLEDDR 840
QY 841 LVHRLAARVAVKSPNHVYITDEGLARLDDIDETEHADGKVPYIMMALESTLRRFT 900
Db 841 LVHRLAARVAVKSPNHVYITDEGLARLDDIDETEHADGKVPYIMMALESTLRRFT 900
QY 901 HOSDWSYGVYVWELMTFGAKPYDGPAREIPDLLEKGERLPORICTIDYIMVWCM 960
Db 901 HOSDWSYGVYVWELMTFGAKPYDGPAREIPDLLEKGERLPORICTIDYIMVWCM 960
QY 961 IDSECRPREBELVSEFSRMAADPQREYVIONEDLGPAFLDSTFYRSLLEDDMDGLVDA 1020
Db 961 IDSECRPREBELVSEFSRMAADPQREYVIONEDLGPAFLDSTFYRSLLEDDMDGLVDA 1020
QY 1021 EBYLVPOQGFCCPDPAAGAGMHHNR 1048
Db 1021 EBYLVPOQGFCCPDPAAGAGMHHNR 1048

```

```

RESULT 8
US-09-854-356-3
; Sequence 3, Application US/09854356
; Patent No. US2002017567A1
; GENERAL INFORMATION:
; APPLICANT: Cheever, Martin A.
; APPLICANT: Cheever, Dirk
; APPLICANT: Corixa Corporation
; APPLICANT: SmithKline Beecham Biologicals S. A.
; TITLE OF INVENTION: HER-2/neu Fusion Proteins
; FILE REFERENCE: 014058-009810PC
; CURRENT APPLICATION NUMBER: US/09/854,356
; CURRENT FILING DATE: 2001-05-09
; PRIOR APPLICATION NUMBER: US 09/493,480
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: US 60/117,976
; PRIOR FILING DATE: 1999-01-29

```

```

; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 3
; LENGTH: 653
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: extracellular domain (ECD) of human HER-2/neu
US-09-854-356-3
Query Match          91.8%; Score 3628; DB 9; Length 653;
Best Local Similarity 100.0%; Pred. No. 1,4e-248;
Matches 653; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 MELAALCRWGLLALLPPGAASPTVCTGTDMKRLRPSPTHLDMLRHLVQGGVQVGNL 60
Db 1 MELAALCRWGLLALLPPGAASPTVCTGTDMKRLRPSPTHLDMLRHLVQGGVQVGNL 60
QY 61 ELTYLPTNASLSFLQDIQEVQYVLIANNOVROYPLQRLRIVGTQLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIQEVQYVLIANNOVROYPLQRLRIVGTQLFEDNYALAVLDNG 120
QY 121 DPLNNTPTVVGASPGGLREIQLSLPTEILKGVLIQRPOLCTQDTLWKDIFHKNNQLA 180
Db 121 DPLNNTPTVVGASPGGLREIQLSLPTEILKGVLIQRPOLCTQDTLWKDIFHKNNQLA 180
QY 181 LTLIDTNRSRACHPCSPMGKSRGWGESSSDCSLPTVCAGCARGKGLPTDCHEQC 240
Db 181 LTLIDTNRSRACHPCSPMGKSRGWGESSSDCSLPTVCAGCARGKGLPTDCHEQC 240
QY 241 AAGCTGKHSDDCLACLFHNSGICELCPALVYNTDTFESMPNPEGRYTFGASCYTACP 300
Db 241 AAGCTGKHSDDCLACLFHNSGICELCPALVYNTDTFESMPNPEGRYTFGASCYTACP 300
QY 301 YNTLSTVGSCTLVCPPLHNOEVAEDGTORCEKSKPCARVCYGLGMEHLREVAAYTSAN 360
Db 301 YNTLSTVGSCTLVCPPLHNOEVAEDGTORCEKSKPCARVCYGLGMEHLREVAAYTSAN 360
QY 361 IOEFAGCKITFGSLAFIPESFDGDPASNTAPLOPEOLQVEETELETGYLYISAMPDLP 420
Db 361 IOEFAGCKITFGSLAFIPESFDGDPASNTAPLOPEOLQVEETELETGYLYISAMPDLP 420
QY 421 DLSVFQMLQYIRGRILHNGAYSLTLQGLISWLGRLSRLGSLALIHNTHLCEVHTV 480
Db 421 DLSVFQMLQYIRGRILHNGAYSLTLQGLISWLGRLSRLGSLALIHNTHLCEVHTV 480
QY 481 PMDOLFNPQALHTANRREDECVGEGSLACHOLCARGHCGMPGPTQCVNCSOFLRGQEC 540
Db 481 PMDOLFNPQALHTANRREDECVGEGSLACHOLCARGHCGMPGPTQCVNCSOFLRGQEC 540
QY 541 VEECRVLQGLPREVYNARHCLPCHPECOPONGSVTCGFPADQCVACAHYKDPFCVARC 600
Db 541 VEECRVLQGLPREVYNARHCLPCHPECOPONGSVTCGFPADQCVACAHYKDPFCVARC 600
QY 601 PSQVKPDLSTYMPIMKFPDEEGACQPCPINCTHSCVDLDKGCRAEQASPLTS 653
Db 601 PSQVKPDLSTYMPIMKFPDEEGACQPCPINCTHSCVDLDKGCRAEQASPLTS 653

```

```

RESULT 9
US-09-921-161-1
; Sequence 1, Application US/09921161
; Patent No. US20020090662A1
; GENERAL INFORMATION:
; APPLICANT: Ralph, Peter
; APPLICANT: ANALYTICAL METHOD
; FILE REFERENCE: GENENT, 066A
; CURRENT APPLICATION NUMBER: US/09/921,161
; CURRENT FILING DATE: 2001-08-01
; PRIOR APPLICATION NUMBER: 60/225,433
; PRIOR FILING DATE: 2000-08-15
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: FastSeq for Windows Version 4.0

```

```

; SEQ ID NO 1
; LENGTH: 645
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-921-161-1

```

```

Query Match      90.8%; Score 3590; DB 10; Length 645;
Best Local Similarity 100.0%; Pred. No. 6.8e-246;
Matches 645; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 MELAALCRMGGLLALLPAGASTOVCTGDMKRLPASPETHIDMLRHLHYOGCQVVGML 60
DB 1 MELAALCRMGGLLALLPAGASTOVCTGDMKRLPASPETHIDMLRHLHYOGCQVVGML 60
OY 61 ELTYLPTNASLSFLDIOEQVGVLIAHNOVROPVLRIRYRGTOLEFEDNTALAVLDNG 120
DB 61 ELTYLPTNASLSFLDIOEQVGVLIAHNOVROPVLRIRYRGTOLEFEDNTALAVLDNG 120
OY 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILMKDIFHRKNOLA 180
DB 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILMKDIFHRKNOLA 180
OY 181 LTLIDNRSRACHPCSPMKGSKRCWGESSEDCQSLRTFYCAGGACARCKGRLPTDCCHQC 240
DB 181 LTLIDNRSRACHPCSPMKGSKRCWGESSEDCQSLRTFYCAGGACARCKGRLPTDCCHQC 240
OY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMNPREGRYTFGASCVTAC 300
DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMNPREGRYTFGASCVTAC 300
OY 301 YNLTSDVGSCTLYCPGLHNOEYTAEDGTORCEKSKPCARVCYGLGMEHLREVRATYSAN 360
DB 301 YNLTSDVGSCTLYCPGLHNOEYTAEDGTORCEKSKPCARVCYGLGMEHLREVRATYSAN 360
OY 361 IOEFAGCKKIFGSLAFPRSPDGPASNTAPLOPEQLQVFEITLITGLYISAMPDLP 420
DB 361 IOEFAGCKKIFGSLAFPRSPDGPASNTAPLOPEQLQVFEITLITGLYISAMPDLP 420
OY 421 DLSVFNQNLQVIRGRIHNGAYSILTQGLISWLGRLSRELSGGLAIHHNTHLCFVHT 480
DB 421 DLSVFNQNLQVIRGRIHNGAYSILTQGLISWLGRLSRELSGGLAIHHNTHLCFVHT 480
OY 481 PMDOLFNRPHOALLHTANRPEDCEVGEGLAQHQLCARHCWGPPTQVCNCSOFLSGQHC 540
DB 481 PMDOLFNRPHOALLHTANRPEDCEVGEGLAQHQLCARHCWGPPTQVCNCSOFLSGQHC 540
OY 541 VEBCRYLQGLPREYVNAHRCPCHEPCOPQNSVTCFEPADQCVACAHYKDPPECVAC 600
DB 541 VEBCRYLQGLPREYVNAHRCPCHEPCOPQNSVTCFEPADQCVACAHYKDPPECVAC 600
OY 601 PGVVPDLSTYMPIMKFPDEEGACQPCPINCTHSCVDLDDKCGPAE 645
DB 601 PGVVPDLSTYMPIMKFPDEEGACQPCPINCTHSCVDLDDKCGPAE 645

```

```

RESULT 10
US-09-854-356-2

```

```

; Sequence 2, Application US/09854356
; Patent No. US20020177567A1
; GENERAL INFORMATION:
; APPLICANT: Cheever, Martin A.
; APPLICANT: Ghayesen, Dirk
; APPLICANT: Corixa Corporation
; APPLICANT: SmithKline Beecham Biologicals S. A.
; TITLE OF INVENTION: HER-2/neu Fusion Proteins
; FILE REFERENCE: 014058-009810PC
; CURRENT APPLICATION NUMBER: US/09/854,356
; PRIOR FILING DATE: 2001-05-09
; PRIOR APPLICATION NUMBER: US 09/493,480
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: US 60/117,976
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 26

```

```

; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1256
; TYPE: PRT
; ORGANISM: Rattus sp.
; FEATURE:
; OTHER INFORMATION: rat HER-2/neu protein
; NAME/KEY: DOMAIN
; LOCATION: (1)..(654)
; OTHER INFORMATION: extracellular domain (ECD)
; NAME/KEY: DOMAIN
; LOCATION: (677)..(1256)
; OTHER INFORMATION: intracellular domain (ICD)
; NAME/KEY: DOMAIN
; LOCATION: (721)..(998)
; OTHER INFORMATION: kinase domain (KD)
; NAME/KEY: DOMAIN
; LOCATION: (991)..(1049)
; OTHER INFORMATION: phosphorylation domain (PD)
; OTHER INFORMATION: fragment of the phosphorylation domain, preferred
; OTHER INFORMATION: portion (delta PD)
US-09-854-356-2

```

```

Query Match      81.2%; Score 3209.5; DB 9; Length 1256;
Best Local Similarity 58.0%; Pred. No. 1.1e-218;
Matches 608; Conservative 34; Mismatches 70; Indels 337; Gaps 2;

```

```

OY 1 MELAALCRMGGLLALLPAGASTOVCTGDMKRLPASPETHIDMLRHLHYOGCQVVGML 60
DB 1 MELAALCRMGGLLALLPAGASTOVCTGDMKRLPASPETHIDMLRHLHYOGCQVVGML 60
OY 61 ELTYLPTNASLSFLDIOEQVGVLIAHNOVROPVLRIRYRGTOLEFEDNTALAVLDNG 120
DB 61 ELTYLPTNASLSFLDIOEQVGVLIAHNOVROPVLRIRYRGTOLEFEDNTALAVLDNG 120
OY 121 DPLNNTPTVT-GASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILMKDIFHRKNOL 179
DB 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILMKDIFHRKNOL 180
OY 180 ALTLIDNRSRACHPCSPMKGSKRCWGESSEDCQSLRTFYCAGGACARCKGRLPTDCCHQ 239
DB 180 ALTLIDNRSRACHPCSPMKGSKRCWGESSEDCQSLRTFYCAGGACARCKGRLPTDCCHQ 239
OY 240 CAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMNPREGRYTFGASCVTAC 299
DB 240 CAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMNPREGRYTFGASCVTAC 299
OY 300 PNYLTSDVGSCTLYCPGLHNOEYTAEDGTORCEKSKPCARVCYGLGMEHLREVRATYSAN 359
DB 300 PNYLTSDVGSCTLYCPGLHNOEYTAEDGTORCEKSKPCARVCYGLGMEHLREVRATYSAN 360
OY 360 NIOEFAGCKKIFGSLAFPRSPDGPASNTAPLOPEQLQVFEITLITGLYISAMPDLP 419
DB 360 NIOEFAGCKKIFGSLAFPRSPDGPASNTAPLOPEQLQVFEITLITGLYISAMPDLP 419
OY 420 PDLVSFQNLQVIRGRIHNGAYSILTQGLISWLGRLSRELSGGLAIHHNTHLCFVHT 479
DB 420 PDLVSFQNLQVIRGRIHNGAYSILTQGLISWLGRLSRELSGGLAIHHNTHLCFVHT 480
OY 480 VPMDOLEFRPHOALLHTANRPEDCEVGEGLAQHQLCARHCWGPPTQVCNCSOFLSGQHC 539
DB 480 VPMDOLEFRPHOALLHTANRPEDCEVGEGLAQHQLCARHCWGPPTQVCNCSOFLSGQHC 540
OY 540 CVEBCRYLQGLPREYVNAHRCPCHEPCOPQNSVTCFEPADQCVACAHYKDPPECVAC 599
DB 540 CVEBCRYLQGLPREYVNAHRCPCHEPCOPQNSVTCFEPADQCVACAHYKDPPECVAC 599
OY 600 CGSVKVPDLSTYMPIMKFPDEEGACQPCPINCTHSCVDLDDKCGPAEQRASPLT----- 652
DB 600 CGSVKVPDLSTYMPIMKFPDEEGACQPCPINCTHSCVDLDDKCGPAEQRASPLT----- 660

```

QY 653 ----- 652  
Db 661 GVLLELIVVVGILIKRRROKIRKYMRLLOETELVEPLTPSGAMPNOAMRIKETE 720  
QY 653 ----- 652  
Db 721 LRKVKLGSGAFGYKGIWIPDGENKIPVAIKVIRENTSPKANKETLDEAYVAGVGS 780  
QY 653 ----- 652  
Db 781 PYVSRLLGICLTSTVOLYQMLPYGCLLDHYREHGRGLSGODLLMVCQIAKMSYLEDV 840  
QY 653 ----- 652  
Db 841 RLVRHDLAARNLVKSPNHVKITDGLARLDIDETEXHADGKVPKMMALESIILRRRF 900  
QY 653 ----- 652  
Db 901 THOSDWSYGVYVWELMTFGAKPYDGIIPAREIPDLLEKGERLPDPPICTIDVYIMVYKWC 960  
QY 653 ----- 652  
Db 961 MIDSECRPRRELVSERSMARDPQRFVYIONEDLGSPSPMDSTFYRSLLEDDMGDLVD 1020  
QY 684 AEETLVPOQGFPCPDPAAGAGMHHRR 712  
Db 1021 AEETLVPOQGFSPDPPTGTGTAHRRHR 1049

RESULT 11  
US-09-870-759-118  
; Sequence 118, Application US/09870759  
; Patent No. US20020177551A1  
; GENERAL INFORMATION:  
; APPLICANT: TERMAN, David S  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT OF NEOPLASTIC DISEASE  
; FILE REFERENCE: 870759  
; CURRENT APPLICATION NUMBER: US/09/870, 759  
; CURRENT FILING DATE: 2002-01-14  
; PRIOR APPLICATION NUMBER: US 60/208,128  
; PRIOR FILING DATE: 2000-05-30  
; NUMBER OF SEQ ID NOS: 166  
; SOFTWARE: Patent version 3.1  
; SEQ ID NO 118  
; LENGTH: 1260  
; TYPE: PRT  
; ORGANISM: Rattus norvegicus  
; US-09-870-759-118

Query Match 81.1%; Score 3205; DB 9; Length 1260;  
Best Local Similarity 57.9%; Pred. No. 2,36-218;  
Matches 608; Conservative 34; Mismatches 70; Indels 338; Gaps 3;

QY 1 MELALCRWGLLLALPPGASTOYCTGDMKRLPASPETHLMDLRHLHYGGCOVYVGNL 60  
Db 4 MELAAWCRWGLLLALPPGASTOYCTGDMKRLPASPETHLMDLRHLHYGGCOVYVGNL 63  
QY 61 ELTVPNASLFLQDIOEVGYVLAHNOYRQYRQRLRIVRGOLFEDNYVALAVLDNG 120  
Db 64 ELTVPNASLFLQDIOEVGYVLAHNOYRQYRQRLRIVRGOLFEDNYVALAVLDNR 123  
QY 121 DPLNNTPTVT-GASPGGLRELQSLTEILKGVLIQNRPOLCYODTILMIDIFKNNQL 179  
Db 124 DPGONVASTPGRPEGLRELQSLTEILKGVLIQNRPOLCYODTILMIDIFKNNQL 183  
QY 180 ALTIDINRSACHPCSPCKSGKRWGESSDQSLRTTVCAGGACRGKPLPDDCHEQ 239  
Db 184 APVIDIDINRSACHPCSPCKSGKRWGESSDQSLRTTVCAGGACRGKPLPDDCHEQ 243  
QY 240 CAAGCTGKHSDDLACLFHNSGICELHCPALVYNTDPEFSPMNPBGRYTFGASCVYAC 299  
Db 244 CAAGCTGKHSDDLACLFHNSGICELHCPALVYNTDPEFSPMNPBGRYTFGASCVYTC 303

QY 300 PYNLTSDVSGCTLVCPRLHNOETAEADGTORCEKSCPKACAVCYGLGMEHLREVAYTSA 359  
Db 304 PYNLTSTEVSGCTLVCPRLHNOETAEADGTORCEKSCPKACAVCYGLGMEHLRGARATISD 363  
QY 360 NIOEPAGCKRIEGLAPLESFDDPASNTAPLQPEOLQVETLEITGYLYISAMPDSL 419  
Db 364 NVOEPAGCKRIEGLAPLESFDDPASNTAPLQPEOLQVETLEITGYLYISAMPDSL 423  
QY 420 PDLVFPQMLQYIRGRILHNGAVSLTGLGISTWLGRLSLRELISGLALIHNTHLCEVHT 479  
Db 424 RDLVFPQMLRIRGRILHNGAVSLTGLGISTWLGRLSLRELISGLALIHNTHLCEVHT 483  
QY 480 VPMQDLFRNPQALLHTANRDEDE-CVGEGLACIQLCARGCKWGPPTQCVNCSOFLRQ 538  
Db 484 VPMQDLFRNPQALLHTANRDEDE-CVGEGLACIQLCARGCKWGPPTQCVNCSOFLRQ 543  
QY 539 ECVEECRVLQGLPREYVYARHCLPCHPCOPONSVCYFGEADQCVACAHYKDRPCVA 598  
Db 544 ECVEECRVKGLPREYVYARHCLPCHPCOPONSVCYFGEADQCVACAHYKDRPCVA 603  
QY 599 RCPGCVKPDLSYMPIMKPPDEEGACQPCPINCTHSCVDLDKCGPAEQRASPLT----- 652  
Db 604 RCPGCVKPDLSYMPIMKPPDEEGACQPCPINCTHSCVDLDKCGPAEQRASPLTATV 663  
QY 653 ----- 652  
Db 664 EGVLLFLIVVVGILIKRRROKIRKYMRLLOETELVEPLTPSGAMPNOAMRIKET 723  
QY 653 ----- 652  
Db 724 ELRKVKLGSGAFGYKGIWIPDGENKIPVAIKVIRENTSPKANKETLDEAYVAGVGS 783  
QY 653 ----- 652  
Db 784 SPYVSRLLGICLTSTVOLYQMLPYGCLLDHYREHGRGLSGODLLMVCQIAKMSYLEDV 843  
QY 653 ----- 652  
Db 844 RLVRHDLAARNLVKSPNHVKITDGLARLDIDETEXHADGKVPKMMALESIILRR 903  
QY 653 ----- 652  
Db 904 THOSDWSYGVYVWELMTFGAKPYDGIIPAREIPDLLEKGERLPDPPICTIDVYIMVYK 963  
QY 653 ----- 652  
Db 964 WMIDSECRPRRELVSERSMARDPQRFVYIONEDLGSPSPMDSTFYRSLLEDDMGDLV 1023  
QY 683 DAETLVPOQGFPCPDPAAGAGMHHRR 712  
Db 1024 DAETLVPOQGFSPDPPTGTGTAHRRHR 1053

RESULT 12  
US-09-854-356-14  
; Sequence 14, Application US/09854356  
; Patent No. US20020177567A1  
; GENERAL INFORMATION:  
; APPLICANT: Cheever, Martin A.  
; APPLICANT: Gheysen, Dirk  
; APPLICANT: Corixa Corporation  
; APPLICANT: SmithKline Beecham Biologicals S. A.  
; TITLE OF INVENTION: HER-2/neu Fusion Proteins  
; FILE REFERENCE: 014058-009810PC  
; CURRENT APPLICATION NUMBER: US/09/854, 356  
; CURRENT FILING DATE: 2001-05-09  
; PRIOR APPLICATION NUMBER: US 09/493,480  
; PRIOR FILING DATE: 2000-01-28  
; PRIOR APPLICATION NUMBER: US 60/117,976  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: Patent version 2.1  
; SEQ ID NO 14

LENGTH: 1256  
TYPE: prt  
ORGANISM: Mus sp.  
FEATURE:  
OTHER INFORMATION: mouse HER-2/neu protein  
US-09-854-356-14

Query Match 80.7%; Score 3189.5; DB 9; Length 1256;  
Best Local Similarity 57.8%; Pred. No. 2.9e-217;  
Matches 606; Conservative 36; Mismatches 70; Indels 337; Gaps 2;

OY 1 MELALICRMGLLALLPGAASTGYCTGDMKRLPASPETHLDMRLHYOCQYVQGNL 60  
DB 1 MELAMCMGFLALLSGAGTGYCTGDMKRLPASPETHLDMRLHYOCQYVQGNL 60  
OY 61 ELTYLPTNASSLFLODIEVOGYVLIHNOVRQVPLQRLIRIVRGTLQFEDNTALAVLNG 120  
DB 61 ELTYLPTNASSLFLODIEVOGYVLIHNOVRQVPLQRLIRIVRGTLQFEDNTALAVLNR 120  
OY 121 DPLNN-TTPVNGASPGRLRELOLRSLTELKGVLIQNRPOCYODTILMKDIFHKNNQL 179  
DB 121 DPLDVTTPAARGTRPEGRLRELOLRSLTELKGVLIQNRPOCYODMYLMKVLRKNNQL 180  
OY 180 ALTLIDTNRSRACHPSPCKSGRCMGSESDCQSLTRTVACGACGARKGPLPTDCCHQ 239  
DB 181 APVDMDTNRSRACHPSPCKSLTRTVACGACGARKGPLPTDCCHQ 240  
OY 240 CAAGCTGPKHSDCLACLFHNHSGICELHCPALVTYNTDTFESMPNDEGRYTFGASCVTAC 299  
DB 241 CAAGCTGPKHSDCLACLFHNHSGICELHCPALVTYNTDTFESMLNDEGRYTFGASCVTTC 300  
OY 300 PYNLTSTVGSCTLVCPPLHNOVTAEDGTQRECKSKPCARCYGLGMEHLREVAAYTSA 359  
DB 301 PYNLTSTVGSCTLVCPPLHNOVTAEDGTQRECKSKPCARCYGLGMEHLREVAAYTSD 360  
OY 360 NIOEFAGCKKIFGSLAFPLPESFDGDPASNTAPLOPEOLOVFTLEITGYLISAMPDSL 419  
DB 361 NIOEFAGCKKIFGSLAFPLPESFDGDPASNTAPLOPEOLOVFTLEITGYLISAMPSEF 420  
OY 420 PDLSTFQNLQVIRGRILHNGAYSLTLQIGISWLGRLSLRELGSGLALIHNTHLCEVHT 479  
DB 421 PDLSTFQNLQVIRGRILHNGAYSLTLQIGISWLGRLSLRELGSGLALIHNTHLCEVHT 480  
OY 480 VPMDOLFNPFOALHTFANRPDECEVCGSLACHOACGHCMPGPTOCVNCISOFLRGDE 539  
DB 481 VPMDOLFNPFOALHTFANRPDECEVCGSLACHOACGHCMPGPTOCVNCISOFLRGDE 540  
OY 540 CVEECRVLOGLPREYVNAHNLPCHPRECOPONGSVTCFEPADOCVACAHYKDPFCVYAR 599  
DB 541 CVEECRVLOGLPREYVNAHNLPCHPRECOPONGSVTCFEPADOCVACAHYKDPFCVYAR 600  
OY 600 CPSSGKRPDLSTYPIKWFDEBAGACOPPCPNCTHSCVDLDDKCPADEQRASPLT----- 652  
DB 601 CPSSGKRPDLSTYPIKWFDEBAGACOPPCPNCTHSCVDLDDKCPADEQRASPLT----- 652  
OY 653 ----- 652  
DB 661 GYLLELLIIVVIGILIKRRRQKIRKTYTMRLLQETELVEPLTPSGAVPNQAOAIRLKETE 720  
OY 653 ----- 652  
DB 721 LKRLKLVSGAGTYVKGIMIPDGENVKIPYAIKVLRENTSPKANKEILDEAYVMAAGVS 780  
OY 653 ----- 652  
DB 781 PYVSHLIGICLSTVQLVTLMPYGLLDHVRHNRGLSGODLNMWCVOIAKMSYLEEV 840  
OY 653 ----- 652  
DB 841 RLVHNDLARNVLYKSPMNVKITPDEGLARLLDIDETEVHADGKVP IKMMALLESILRRRF 900  
OY 653 ----- 652

DB 901 THQSDVWSGYVTWMLMTFGAKPYDGIAPAREIPDLLEKGERLPQPPICITIDVYIMVMKCM 960  
OY 653 -----SQNEDLGPAISPDLSTYRSLLEDDMGDLVD 683  
DB 961 MIDSECRPRFRELVESEFSMARDPQRFVYIQONEDLGPSPMSTYRSLLEDDMGDLVD 1020  
OY 684 AEEYLVPOOGFPCPDPAAGAGMVHRRH 712  
DB 1021 AEEYLVPOOGFPCPDPAAGAGMVHRRH 1049

RESULT 13  
US-09-854-356-8  
Sequence 8, Application US/09854356  
Patent No. US2002017567A1  
GENERAL INFORMATION:  
APPLICANT: Cheever, Martin A.  
APPLICANT: Cheyssen, Dirk  
APPLICANT: Corixa Corporation  
APPLICANT: SmithKline Beecham Biologicals S. A.  
TITLE OF INVENTION: HER-2/neu Fusion Proteins  
FILE REFERENCE: 014058-009810PC  
CURRENT APPLICATION NUMBER: US/09/854,356  
CURRENT FILING DATE: 2001-05-09  
PRIOR APPLICATION NUMBER: US 09/493,480  
PRIOR FILING DATE: 2000-01-28  
PRIOR APPLICATION NUMBER: US 60/117,976  
PRIOR FILING DATE: 1999-01-29  
NUMBER OF SEQ ID NOS: 26  
SOFTWARE: Patentln Ver. 2.1  
SEQ ID NO 8

LENGTH: 654  
TYPE: prt  
ORGANISM: Rattus sp.  
FEATURE:  
OTHER INFORMATION: extracellular domain (ECD) of rat HER-2/neu  
US-09-854-356-8

Query Match 78.7%; Score 3110.5; DB 9; Length 654;  
Best Local Similarity 85.5%; Pred. No. 4.8e-212;  
Matches 558; Conservative 32; Mismatches 62; Indels 1; Gaps 1;

OY 1 MELALICRMGLLALLPGAASTGYCTGDMKRLPASPETHLDMRLHYOCQYVQGNL 60  
DB 1 MELAMCMGFLALLPGAASTGYCTGDMKRLPASPETHLDMRLHYOCQYVQGNL 60  
OY 61 ELTYLPTNASSLFLODIEVOGYVLIHNOVRQVPLQRLIRIVRGTLQFEDNTALAVLNG 120  
DB 61 ELTYLPTNASSLFLODIEVOGYVLIHNOVRQVPLQRLIRIVRGTLQFEDNTALAVLNR 120  
OY 121 DPLNN-TTPVNGASPGRLRELOLRSLTELKGVLIQNRPOCYODTILMKDIFHKNNQL 179  
DB 121 DPLDVTTPAARGTRPEGRLRELOLRSLTELKGVLIQNRPOCYODMYLMKVLRKNNQL 180  
OY 180 ALTLIDTNRSRACHPSPCKSGRCMGSESDCQSLTRTVACGACGARKGPLPTDCCHQ 239  
DB 181 APVDMDTNRSRACHPSPCKSLTRTVACGACGARKGPLPTDCCHQ 240  
OY 240 CAAGCTGPKHSDCLACLFHNHSGICELHCPALVTYNTDTFESMPNDEGRYTFGASCVTAC 299  
DB 241 CAAGCTGPKHSDCLACLFHNHSGICELHCPALVTYNTDTFESMLNDEGRYTFGASCVTTC 300  
OY 300 PYNLTSTVGSCTLVCPPLHNOVTAEDGTQRECKSKPCARCYGLGMEHLREVAAYTSA 359  
DB 301 PYNLTSTVGSCTLVCPPLHNOVTAEDGTQRECKSKPCARCYGLGMEHLREVAAYTSD 360  
OY 360 NIOEFAGCKKIFGSLAFPLPESFDGDPASNTAPLOPEOLOVFTLEITGYLISAMPDSL 419  
DB 361 NIOEFAGCKKIFGSLAFPLPESFDGDPASNTAPLOPEOLOVFTLEITGYLISAMPDSL 420  
OY 420 PDLSTFQNLQVIRGRILHNGAYSLTLQIGISWLGRLSLRELGSGLALIHNTHLCEVHT 479  
DB 421 PDLSTFQNLQVIRGRILHNGAYSLTLQIGISWLGRLSLRELGSGLALIHNTHLCEVHT 480





GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 13, 2003, 14:44:35 ; Search time 32.7407 Seconds  
(without alignments)  
4480.838 Million cell updates/sec

Title: US-09-854-356-7  
Perfect score: 3954  
Sequence: 1 MELALACRMGLLALLPFGA.....GFCCPDPAAGAGVHHRRH 712

Scoring table: ELOSUM62  
Gapop 10.0 , Gapept 0.5

Searched: 671580 seqs, 206047115 residues  
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_21:\*  
1: sp.\_archaea:\*  
2: sp.\_bacteria:\*  
3: sp.\_fungi:\*  
4: sp.\_human:\*  
5: sp.\_invertebrate:\*  
6: sp.\_mammal:\*  
7: sp.\_mhc:\*  
8: sp.\_organelle:\*  
9: sp.\_phage:\*  
10: sp.\_plant:\*  
11: sp.\_rodent:\*  
12: sp.\_virus:\*  
13: sp.\_vertebrate:\*  
14: sp.\_unclassified:\*  
15: sp.\_virus:\*  
16: sp.\_bacteriap:\*  
17: sp.\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3446.5	87.2	1259	6	018735
2	1874.5	47.4	419	4	090K79
3	1540	38.9	1209	11	090X70
4	1533.5	38.8	655	11	090WFS
5	1533.5	38.8	1210	11	090EP98
6	1517.5	38.4	643	11	090ERV6
7	1348.5	34.1	1328	13	P79754
8	1346.5	34.1	1165	13	09YH40
9	1156.5	29.2	1137	13	090W6F6
10	1155	29.2	527	13	090836
11	1030.5	26.1	1433	5	09B1H9
12	1028.5	26.0	478	11	090ESF0
13	971.5	24.6	599	13	090PSH2
14	887	22.4	176	11	0923VS
15	734	18.6	331	4	090BD7
16	723	18.3	149	6	090B66

17	656	16.6	1368	5	023821	023821 caenorhabdi
18	621	15.7	1717	5	026566	026566 schistosoma
19	620.5	15.7	1193	5	09Y1X8	09Y1X8 ephyattia f
20	406.5	10.3	1362	13	09PY24	09PY24 xenopus lae
21	404.5	10.2	150	6	09BG64	09BG64 oryctolagus
22	395	10.0	151	6	09BG65	09BG65 oryctolagus
23	366.5	9.3	1368	5	090W85	090W85 paralicthy
24	366.5	9.3	1472	13	090U88	090U88 bombyx mori
25	363.5	9.2	366	5	026569	026569 schistosoma
26	362	9.2	1369	13	080W86	080W86 paralicthy
27	359	9.1	1358	13	073798	073798 xenopus lae
28	350.5	8.9	136	11	090PK1	090PK1 mus musculu
29	345.5	8.7	1418	13	093457	093457 scophthalmu
30	340.5	8.6	1671	5	090NV5	090NV5 biophalar
31	339.5	8.6	410	11	063720	063720 rattus norv
32	338	8.5	412	4	08WYV0	08WYV0 homo sapien
33	331	8.4	334	5	026567	026567 schistosoma
34	331	8.4	342	5	026568	026568 schistosoma
35	328	8.3	149	11	060494	060494 cavia sp. e
36	326.5	8.3	1245	13	09YGH8	09YGH8 scophthalmu
37	325	8.2	1418	13	080W83	080W83 paralicthy
38	322	8.1	469	11	063721	063721 rattus norv
39	316	8.0	1371	11	090VW4	090VW4 rattus norv
40	315	8.0	1412	13	080W84	080W84 paralicthy
41	307	7.8	2144	5	090V94	090V94 drosophila
42	302.5	7.7	1749	5	08T0M6	08T0M6 echinococcu
43	296	7.5	946	5	090V04	090V04 drosophila
44	284.5	7.2	116	6	090B67	090B67 oryctolagus
45	281	7.1	367	11	08R2X1	08R2X1 mus musculu

## ALIGNMENTS

RESULT 1  
ID 018735 PRELIMINARY; PRT; 1259 AA.  
AC 018735;  
DT 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Erbb-2.  
OS Canis familiaris (Dog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
OX NCBI\_TaxID=9615;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Yokota H.;  
RT "CDNA cloning of erbb-2 from canine mammary gland."  
RL Submitted (0CT-1997) to the EMBL/GenBank/DBS databases.  
DR EMBL; AB008451; BAA23127.1; -.  
DR HSSP; P1362; IFGK.  
DR InterPro; IPR002048; EF-hand.  
DR InterPro; IPR000494; EGFR\_L\_domain.  
DR InterPro; IPR000719; Euk\_Pkinase.  
DR InterPro; IPR002174; Furin-like.  
DR InterPro; IPR001245; Tyr\_Pkinase.  
DR InterPro; IPR004019; YLP\_motif.  
DR Pfam; PF00757; Furin-like; 1.  
DR Pfam; PF00069; Pkinase; 1.  
DR Pfam; PF01030; Recep\_L\_domain; 2.  
DR Pfam; PF02757; YLP; 2.  
DR ProDom; PD000001; Euk\_Pkinase; 1.  
DR SMART; SM00219; TYKc; 1.  
DR SMART; SM00219; TYKc; 1.  
DR PROSITE; PS00018; EF\_HAND; UNKNOWN; 1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS00011; PROTEIN\_KINASE\_DOM; 1.  
DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
KW ATP-binding; Transferase; Tyrosine-protein kinase.  
SQ SBOUNCE 1259 AA; 137989 MW; E37364D9C4ACD46 CRC64;



DT 01-MAY-2000 (Tremblrel. 13, Created)  
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE Epidermal growth factor receptor.  
 GN EGFR.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 NCBI\_TaxID=10116;  
 RX MEDLINE=90258888; PubMed=2342466;  
 RA Petch L.A., Harris J., Raymond V.W., Blasband A.J., Lee D.C.,  
 RA Earp H.S.;  
 RT "A truncated, secreted form of the epidermal growth factor receptor is  
 RT encoded by an alternatively spliced transcript in normal rat tissue.";  
 RL Mol. Cell. Biol. 10:2973-2982(1990).  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FISHER; TISSUE=LIVER;  
 RA Petch L.A.;  
 RL Submitted (NOV-1991) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FISHER; TISSUE=LIVER;  
 RA Cuttridge K., Dawson T.L., Earp H.S.;  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; M37394; AAF14008.1; -.  
 DR HSSP; P11362; 1FCR.  
 DR InterPro: IPR000494; EGFR\_L.domain.  
 DR InterPro: IPR000719; Euk.pkinase.  
 DR InterPro: IPR002174; Furin-like.  
 DR InterPro: IPR001245; Tyr.pkinase.  
 DR Pfam; PF00757; Furin-like; 1.  
 DR Pfam; PF01069; pkinase; 1.  
 DR Pfam; PF01030; Recep\_L.domain; 2.  
 DR PRINTS; PR00109; TYRKINASE.  
 DR Prodom; PD000001; Euk\_pkinase; 1.  
 DR SMART; SM00261; F0; 3.  
 DR SMART; SM00219; Tyrc; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
 KW ATP-binding; Receptor; Transferase; Tyrosine-protein kinase.  
 SQ SEQUENCE 1209 AA; 134891 MW; 96FE7F6CC1B7773 CRC64;

Query Match 38.9%; Score 1540; DB 11; Length 1209;  
 Best Local Similarity 43.6%; Pred. No. 6,7e-130;  
 Matches 291; Conservative 101; Mismatches 240; Indels 36; Gaps 9;

QY 3 LAALCWMGLLLALPPGA-ASTOVCTGTDMKRLRPSPTHTLMDLRHLYOGCVQGNTE 61  
 DB 15 LAALCAAG-----GALEEKVVCOGTSNRLTQLTGFEDHFLSLQRMFNCEVVLGNLE 66  
 QY 62 LTVLPNASTLSFLDIOEYGVYLIHNOYROYVLORLRVRGSTOFEDYVALAVLDNDG 121  
 DB 67 ITTVQNNYDLSFLKTIQEVAGYVLIHNTVERLEPDLITIKRNALYENTIAVLASN- 124  
 QY 122 PLNNTPTVYGASPGRLRLSLTEILKGVLIQNRNPOLCYDITLMDIFKRNQAL 181  
 DB 125 -----GYNKGTGRELPNRLQELIGAVRFNSNNPILCNMETIOWNRIV-QDVELSN 175  
 QY 182 TLIDTNRS-RACHPCSPMKSGKSCNCESESDQSLRTVACAGGA-RCKGRLPTDCCHQ 239  
 DB 176 MSMDVORHLTGCPKCDPSPNGSCWGEENCKLRKIIICAQCSRRCRSPSCDCHQ 235  
 QY 240 CAAGCTGPKHSDCLALHFNHSGICELHCPALVYNTDFEESPMPNEGRRTFGASCVTAG 299  
 DB 236 CAAGCTGPKHSDCLVCHRRDEATCKDTCRPLMLYNTTYQMDVNEBGKTSFATCYKAC 295  
 QY 300 PNYVLTSDVGSCTLVCLPLHNOEVTADGTQRCCKSCPCARVCYGLGMEHLREAVRTSA 359

DB 296 PRNYVVTDHSSCVACGPDYEEV-EEDGVSKCKCKDGPCKRVCNGIGEFKDTLSINAT 354  
 QY 360 NIDEPACCKIIPESLAFIPSPFGDPASNTAPLQPPLOVFELEETIGLYISAMPDSL 419  
 DB 355 NIHFYCYTAISGDLHLPLPAFGKDSFTTRPPLDPRELEILTKYKEITGELLQAMPENW 414  
 QY 420 PDSVFNQNVYGRILHNGAYSLTLOGIGISWAGRSLELSGALHNNHTLGFVHT 479  
 DB 415 TDLHAFENLEITGRKTKQKQFSLAVVGLNTISLGRSLAKEISDGVYISGNRLCYANT 474  
 QY 480 VPMDOLEFRNPHQALLTANRDEDECVGEGLACHQLCARHGWGPPTQCVNCSQPLRGE 539  
 DB 475 INMKLIFGPNQKTKIMNNRAEKDKATNNVCPLDSSSECGEPPTDCSCGNVSRGR 534  
 QY 540 CVEBCRYLOGLPREYVNAHRCLEPCHECOPQNSVTCFGEADQVACAHYKDPPCVAR 599  
 DB 535 CVDKCNILKEEPREFENSECIOCHPCDLPQTNNICTGKGPNCIKCAHYVDGPHCVXT 594  
 QY 600 CPESGVKPDLSYMPIMKPEDESGACPCPINCPTSCVDLDKGPAPGRASPLSQKEDIG 659  
 DB 595 CPESGIGENNNTL-VMKFADANNVCHLCHANCITGCGAGPLKGCQPE-----G 641  
 QY 660 PASPLDST 667  
 DB 642 PKIPSIAT 649

RESULT 4  
 ID 09WVF5 PRELIMINARY; PRT; 655 AA.  
 AC 09WVF5;  
 DT 01-NOV-1999 (Tremblrel. 12, Created)  
 DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE Epidermal growth factor receptor (Epidermal growth factor receptor  
 DE Isoform 3).  
 GN EGFR.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=LIVER;  
 RA Reiter J.L., Threadgill D.W., Danielson A.J., Schell C.,  
 RA Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,  
 RA Mahle N.J.;  
 RT "Alternative transcripts from the Human and Mouse EGFR Genes Encode  
 RT Carboxy-Terminal Truncated Receptors";  
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C3H/101, 129/SVJ, AND 129/SVEVYAC;  
 RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,  
 RA Sinclair C.S., Pearisall R.S., Green P.J., Yee D., Lampland A.L.,  
 RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,  
 RA Mahle N.J.;  
 RT "Comparative genomic sequence analysis and isolation of human and  
 RT mouse alternative Egfr transcripts encoding truncated receptor  
 RT isoforms";  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=LIVER;  
 RX MEDLINE=2108560; PubMed=11217851;  
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Aikawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochia H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,

RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldi M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Guenich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Nomberts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
 RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlski S.,  
 RA Hayashizaki Y.;  
 RA "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 DR EMBL; AF124513; AAD4149.1; -;  
 DR EMBL; AF275366; AAG28047.1; -;  
 DR EMBL; AF275364; AAG28047.1; JOINED.  
 DR EMBL; AF275365; AAG28047.1; JOINED.  
 DR EMBL; AK004944; BAB23688.1; -;  
 DR EMBL; AK004883; BAB23641.1; -;  
 DR EMBL; AK004911; BAB23662.1; -;  
 DR MGD; MGI:95294; Egfr.  
 DR InterPro: IPR000494; EGFR\_Ldomain.  
 DR InterPro: IPR002174; Furin-like.  
 DR Pfam; PF00757; Furin-like; 1.  
 DR Pfam; PF01030; Recep\_Ldomain; 2.  
 DR SMART; SM00261; FU; 3.  
 KW Receptor.  
 SO "SEQUENCE 655 AA; 72906 MW; 6B34063B1BC928CB CRC64;

Query Match 38.8%; Score 1533.5; DB 11; Length 655;  
 Best Local Similarity 44.4%; Pred. No. 1.2e-129;  
 Matches 284; Conservative 99; Mismatches 233; Indels 23; Gaps 7;

QY 11 LLLALLPAGA--STVCGTGMKRLPASPEHLDMLRHLYQGCGVQGNLELYPTN 68  
 Db 14 LITLCAAGAGALEEKVKVCGTGNRLTQGTGFEDHFLSLQRMNCEVGLNLEITYVORN 73  
 QY 69 ASLSFLQDIQEVGYVLIAHNOVROPLQRLIVRGTOLEFEDNYALAVLNDGPNLNTTP 128  
 Db 74 YDLSPFLTKTQEVAGVLIALNTVERIPLENQITGNALYENTYALALISN----- 124  
 QY 129 VTGASPGGLRELQLSLEIILKGVLIQRPOLCYQDITLTKD-----FKHKNQIALTLI 184  
 Db 125 -YGTNRRTGLRELPMRNOEILIGAVFSNNPILCNMDTIQWMDIQVNFMSNMSMDL--- 180  
 QY 185 DTRNSRACHPCSPMGKSGRCWSESDCSLTRTVCAGGCA-RCKGPLYTDCCHRCAG 243  
 Db 181 -QSHSSCPKCDPSPNCSCMGGEENCKLTKITCAQCCSHRCGRSPDCCNQCAG 239  
 QY 244 CTGPRHSDCLAGLHNHSGICEIHCALVYNTDTFESMPNPEGRYTFGASCVTACPNY 303  
 Db 240 CTGPRESCLVQKQFQDQATCKDTCRPLMLXNPTTYQMDVNPBGKYSFGATCVKRCPNY 299  
 QY 304 LSTVSGCTVCPRLNNOVETADGTQRCCKSPCARVCYGLGMHLEAVANTANIOE 363  
 Db 300 VVTBHGSCVRCAGPRDYEV-EDDGIRKCKCDGPRKVCNGIGIEFQDTLSINATNIKH 358  
 QY 364 FAGCKKIGSLAFPEPSDFDPSANTAPRLPOLQVFFLEETITGVLYISAMPDLS 423  
 Db 359 FKYCAISGDLHILPVAKKGDSFTTPRLDRELEIKTKITGLTFFLIQAPMDWTIDH 418  
 QY 424 VFONLQVIRGLINGAYSLTLOGISLWLSRLSRLSGSLALHNTLHLCFTVTYWD 483  
 Db 419 AFENLEIRGRKQKQSLAVVGLNITSLGRSLKETSIGRDLVILISGRNLCYANTIMK 478  
 QY 484 QLFNRPHQALLHTANRPDECVGSLAQCLCARHCHSGPFGTQCVNCSQELRGCEVVE 543  
 Db 479 KLFGPRNOKTKIMNRAKDKCAVNVHCNPLCSSSECGMPREDVSCQNSRGRECEYK 538  
 QY 544 CRVLQGLREYVNAHNCCHPECOPOGSGVTCFGRPADOCYACAHYKDPKCVARCPSG 603  
 Db 539 CNILGEPRREVENSECIQCHPECLPQAMNITCTGRPDNCTQCAHYIDGPHCVKTCPRG 598  
 QY 604 VKPDLSTYMPIMKFPDEBAGCOPCPINCTHSCVLDLDCG 642

Db 599 IMGENNTL-VWXYADANNVCHLCHANCYGCAGPGLQCG 636

## RESULT 5

QY 09EP98 PRELIMINARY; PRT; 1210 AA.  
 AC 09EP98;  
 DT 01-MAR-2001 (TRENBLrel. 16, Created)  
 DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)  
 DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)  
 DE Epidermal growth factor receptor isoform 1.  
 GN EGFR.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C3H/101, 129/SVJ, AND 129/SVEVTA;  
 RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,  
 RA Sinclair C.S., Pearall R.S., Green P.J., Yee D., Lampland A.L.,  
 RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,  
 RA Maibale N.J.;  
 RA "Comparative genomic sequence analysis and isolation of human and  
 RT mouse alternative Egfr transcripts encoding truncated receptor  
 RT isoforms.";  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J;  
 RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,  
 RA Schehl C., Pearall R.S., Green P.J., Yee D., Lampland A.L.,  
 RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,  
 RA Maibale N.J.;  
 RA "Comparative genomic sequence analysis and isolation of human and  
 RT mouse alternative Egfr transcripts encoding truncated receptor  
 RT isoforms.";  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 RT Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF275366; AAG28045.1; -;  
 DR EMBL; AF275364; AAG28045.1; JOINED.  
 DR EMBL; AF275365; AAG28045.1; JOINED.  
 DR EMBL; AF275367; AAG24386.1; -;  
 DR HSSP; P11362; IEGK.  
 DR MGD; MGI:95294; Egfr.  
 DR InterPro: IPR000345; Cytochrome\_bbind.  
 DR InterPro: IPR000494; EGFR\_Ldomain.  
 DR InterPro: IPR000719; Euk\_Pkinase.  
 DR InterPro: IPR002174; Furin-like.  
 DR InterPro: IPR002290; Ser\_thr\_Pkinase.  
 DR InterPro: IPR001245; Tyr\_Pkinase.  
 DR Pfam; PF00757; Furin-like; 1.  
 DR Pfam; PF00069; Pkinase; 1.  
 DR Pfam; PF01030; Recep\_Ldomain; 2.  
 DR PRINTS; PR00109; TYRKINASE.  
 DR ProDom; PD000001; Euk\_Pkinase; 1.  
 DR SMART; SM00261; FU; 5.  
 DR SMART; SM00220; S\_TKC; 1.  
 DR SMART; SM00219; TYRK; 1.  
 DR PROSITE; PS00190; CYTOCHROME\_C; UNKNOWN.1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS0011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
 KW ATP-binding; Receptor; Transferase.  
 SO SEQUENCE 1210 AA; 134840 MW; 62CD021C9DE32E18 CRC64;

Query Match 38.8%; Score 1533.5; DB 11; Length 1210;  
 Best Local Similarity 44.4%; Pred. No. 2.6e-129;  
 Matches 284; Conservative 99; Mismatches 233; Indels 23; Gaps 7;

QY 11 LLLALLPAGA--STVCGTGMKRLPASPEHLDMLRHLYQGCGVQGNLELYPTN 68  
 Db 14 LITLCAAGAGALEEKVKVCGTGNRLTQGTGFEDHFLSLQRMNCEVGLNLEITYVORN 73



RA Gellner K., Brenner S.;  
RT "Analysis of 148 kb of genomic DNA around the wnt1 locus of Fugu  
rubripes.";  
RL Genome Res. 9:251-258(1999).  
DR EMBL: AF056116; AAC34391.1; -.  
DR HSSP: P11362; IRGK.  
DR InterPro: IPR000494; EGFR\_L\_domain.  
DR InterPro: IPR000719; Euk\_kinase.  
DR InterPro: IPR002174; Furin-like.  
DR InterPro: IPR001245; Tyr\_kinase.  
DR Pfam: PF00757; Furin-like; 1.  
DR Pfam: PF00069; kinase; 1.  
DR Pfam: PF01030; Recep\_L\_domain; 2.  
DR ProDom: PD000001; Euk\_kinase; 1.  
DR SMART: SM00261; FU; 3.  
DR SMART: SM00219; TYRK; 1.  
DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
DR ATP-binding; transferase.  
KW SEQUENCE 1328 AA; 148613 MW; A333039258B647E9 CRC64;

Query Match 34.1%; Score 1348.5; DB 13; Length 1328;  
Best Local Similarity 43.4%; Pred. No. 1.9e-112;  
Matches 276; Conservative 87; Mismatches 242; Indels 31; Gaps 15;

QY 9 WGLLALPP--GAASD-----VCTGDMKRLPASPETHIDMLRHLVGGCGOVGNLEL 62  
DB 4 WRLIMCVASRLRAASSTQIQAECPTONGSLSTGSOENQVNLNKRKCEIINGNEI 63  
QY 63 TYLPNALSFLQDIOEVOGYVLIANQVNPVLRRLRIVGTQLFEDNYALAVLNDGP 122  
DB 64 TQIESNMDFSLKTRREYTVLIAMNHFOEIPQLGHLVINGNSLYERRFLSVLN-- 120  
QY 123 LNNTPVGAAPGGLREQLSLFPIILKGVLIQNPOLCYQDTILKMDIFHKNQALTL 182  
DB 121 ----YPKG--PSGLNQGLMNLTEILDGVQVILNKKYLRGPVWYWDII--RNDADIE 173  
QY 183 LIDNRSHACHPCSPMCKGRSCWSESSDQSLRTVCAGCG--ARCKGPLPTDCEHQA 241  
DB 174 IQFNGERGVCH--KSC--GNYCWMGKXQDQILITVCAPQCNDSCFETSPDCHIECA 229  
QY 242 AGCGPKASDCLACHFNHSGICELHCPALVYNTDTFESMNPBGRATFGASCVTACPY 301  
DB 230 AGCKGPLDTDFACRLFNDSGACVQCQPTLIYNKQTFOMETNPNAKYQYSGICVSOCP 289  
QY 302 NYLSDVSGCTLVCPCLHNOEYTAEDGTOR--CEKSKPCARVCYIGMHLEVRATVSAN 360  
DB 290 HFV--VDGSSCVSVCPDKMEV--ERGSORQCELCSGLCPKVCEGTGAE---QROTVDSSN 343  
QY 361 IQEFGACKRIGSLAFLPESFSDGPASNTAPLQAPOLQVFTLEETIGYLYISAMPDILP 420  
DB 344 IDSFNCTKIGQSLHFLTGLIGDDFKNVPLDAKKEVFVREITDIILNQSPKELN 403  
QY 421 DLSVFQNIQVIRGLINAGVSLTLOGIGISMLGIRSLRELSGIALIHHNTHLCFVHTV 480  
DB 404 DLSVFSSLTITQGSLSLFRFSIMWNRIPLTLSGLRSLREISDSGVYISQNAHLCYHNTV 463  
QY 481 PMDLQFRPH--QALLHTANRPDECEVGEGLACHOGLCARGHOSGPRPCVNCOSQELRQGE 539  
DB 464 NWTQLFKRSRVRANSLSNRMABECVADGRVCDPLCSGSGWGPEDDCLSCRNSTRHGT 523  
QY 540 CVBECRVLQGLRREVYNAR--CLPCHPCEQPOHNSVTCFGEADQCVACAHYKDPPEVVA 598  
DB 524 CVACGHFNSGIRPREFAGLNGVCAVCHPECKPOTGKASTGCGADCMACSTFRDPRYCMS 583  
QY 599 RCPGCVKPDLSYMPIMKRPDEGACQPCPINCTHSC 634  
DB 584 SCPAGVN--DGEKGLIFKFPNREGHCEPCHONCTGCG 618

RESULT 8  
Q9YH40 PRELIMINARY; PRT; 1165 AA.

AC Q9YH40;  
DT 01-MAY-1999 (TREMBLrel. 10, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Receptor tyrosine kinase proto-oncogene.  
GN XMRK.  
OS Xiphophorus xiphidium.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
OC Acanthomorpha; Acanthopterygii; Percomorph; Atherinomorpha;  
OC Cyprinodontiformes; Poeciliidae; Xiphophorus.  
OX NCBI\_Taxid=8086;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=RIO PURIFICATION;  
EX MEDLINE-98241172; Pubmed-9582016;  
RA Dimitrijevic N., Winkler C., Wellbrock C., Gomez A., Duschl J.,  
RT Altschmid J., Scharl M.;  
RT "Activation of the xmrk proto-oncogene of Xiphophorus by  
overexpression and mutational alterations.";  
RL Oncogene 16:1681-1690(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=RIO PURIFICATION;  
RA Scharl M.;  
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: U53471; AAD10500.2; -.  
DR HSSP: P11362; IRGK.  
DR InterPro: IPR000345; CytC\_heme\_bind.  
DR InterPro: IPR000494; EGFR\_L\_domain.  
DR InterPro: IPR000719; Euk\_kinase.  
DR InterPro: IPR002174; Furin-like.  
DR InterPro: IPR001899; Gram\_pos\_anchor.  
DR InterPro: IPR001245; Tyr\_kinase.  
DR Pfam: PF00757; Furin-like; 1.  
DR Pfam: PF00069; kinase; 1.  
DR Pfam: PF01030; Recep\_L\_domain; 2.  
DR PRINTS: PR00109; TYRKINASE.  
DR ProDom: PD000001; Euk\_kinase; 1.  
DR SMART: SM00261; FU; 3.  
DR SMART: SM00219; TYRK; 1.  
DR PROSITE: PS00190; CYTOCHROME\_C; UNKNOWN\_1.  
DR PROSITE: PS00343; GRAM\_POS\_ANCHORING; UNKNOWN\_1.  
DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 2.  
DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR; 1.  
DR ATP-binding; kinase; transferase; tyrosine-protein kinase.  
KW SEQUENCE 1165 AA; 129614 MW; 7F7EE38D8771A74E CRC64;

Query Match 34.1%; Score 1346.5; DB 13; Length 1165;  
Best Local Similarity 42.1%; Pred. No. 1.9e-112;  
Matches 273; Conservative 91; Mismatches 258; Indels 27; Gaps 12;

QY 1 MELALCRWGLLALPPG-AAST-----QVCTGDMKRLPASPETHIDMLRHLVGGCGOV 55  
DB 4 LELLEL-----LLLLLSIGRCCSTDPDRKVCQGSNQMTM--LDNHYLKKKKMYSGCNV 56  
QY 56 VQGNLELYLPLFNASTSLQDIOEVOGYVLIANQVNPVLRRLRIVGTQLFEDNYALA 115  
DB 57 VLENLEITYTOENDSLQSLQIOEYGVLLIAMEVSTIPLVNLRLRGONLYEGNFTLL 116  
QY 116 VLNDPPLNNTPTVYASPGGLREQLSLFPIILKGVLIQNPOLCYQDTILKMDIFHKN 175  
DB 117 VMSNQR--NPSSP--DYQVGLKQQLSLNLEILISGVKAVSHNPLLCVETIINMMDIYDK 173  
QY 176 NNQALTLIDNRSBACHPCSPMCKGRSCWSESSDQSLRTVCAGCG--ARCKGPLPTD 234  
DB 174 TSNPTMNLIPAFERQCKCDPGCVNGSCWAPRGHCKFKTLCAEQCNRCRGGPKPTD 233  
QY 235 CCHQCAAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDTFESMNPBGRATFGAS 294  
DB 234 CCHQCAAGCTGPRATDCLACRFNDSDGCKDTPCPKATYDVSHQVVDNPNIKYTTGAA 293

Oy		295	CVTACPNVYLSLDVSGSICLVCPHNLNOEVTAEEDTGOREKSKSPCARVCYGLGMEHLREVR	354
			: : :   : :   : :   : :   : :   : :   : :   : :   : :   : :   : :   : :	
Db		294	CVKRCPNRYVTE-GACVRCSSAGMLEVD-ENGRKSRCKPCDGYCPRKCDSDIGISLNTI	351
Oy		355	AVTSANTIOEFAGCKKIGSLAFIPRESFDGPASNTAFLQEPOLQVEPTLEITGYLTISA	414
			: : :   : :   : :   : :   : :   : :   : :   : :   : :   : :   : :   : :	
Db		352	AVNSTINIGSFNCNKINGDIILNNNSFEGDPHYKIGMDPEHLMNLTTVAETIEGYLIWM	411
Oy		415	WPDSLPLDSFPONQLVIRGRILNHGAYS-LTLQGLGISWLGSLNRLSGLLALIHHTH	473
			: : :   : :   : :   : :   : :   : :   : :   : :   : :   : :   : :   : :	
Db		412	WPMNMTSLSYVONELIIRGTFRTERSGRFSFYVVVOYSHLOMGLSLKSEVSAGNYLTKNTPQ	471
Oy		474	LCEFHATVPMDLFENPHOALLHTNANREDDECVGGSLACHOLCARGHCWGPGPQCVCVCSQ	533
			: : :   : :   : :   : :   : :   : :   : :   : :   : :   : :   : :   : :	
Db		472	LRIYSTINMRLETFSEDOQSLEYDART-----ENQTCCNNESSDEGCGWGPPIVCVCLH	524
Oy		534	FLRGOECVECEBRYVLGLPREYVNARHCLPCHPECOPONGSVTGEPPADOCVACAHRKP	593
			: : :   : :   : :   : :   : :   : :   : :   : :   : :   : :   : :   : :	
Db		525	VDRGRCGVASCNLLOGPBREAQVODRGVCVCHQEBELTVGTDSTLVCGPBPANCSCAHRRQG	584
Oy		594	PFCVAPRPSGVKPDLPSTMPYIKPFDEGCACOPCPINTCTHSCVIDIDDKGC	642
			: : :   : :   : :   : :   : :   : :   : :   : :   : :   : :   : :   : :	
Db		585	FOCIIRCPRHGMIGDDFTL-IWKYADKKMGKOCPCQHONCTOGCSGPGLSGC	632
<hr/>				
RESULT 9				
O9w6f6	ID	O9w6f6	PRELIMINARY;	PRT; 1137 AA.
AC	O9w6f6;			
DT	01-NOV-1999 (TREMBLrel. 12, Created)			
DT	01-NOV-1999 (TREMBLrel. 12, Last sequence update)			
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)			
GN	Receptor Tyrosine Kinase (Fragment).			
GN	ERBB4.			
OS	Gallus gallus (Chicken).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;			
OC	Gallus.			
OX	NCBI_TaxID=9031;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	TISSUE=HINDRAIN;			
RX	MEDLINE=99263203; PubMed=10328864;			
RA	Dixon M., Lumsden A.;			
RT	"Distribution of neuregulin-1 (nrq1) and erbB4 transcripts in embryonic chick hindbrain.";			
RL	Mol. Cell. Neurosci. 13:237-258(1999).			
EMBL	AF121963; AACD31764.1; .			
DR	HSP, p11362; IFGK.			
DR	InterPro: IPR000494; EGFR_L_domain.			
DR	InterPro: IPR000719; Euk_pkinase.			
DR	InterPro: IPR002174; Furin-like.			
DR	InterPro: IPR001368; TNFR_c6.			
DR	InterPro: IPR001245; Tyr_Pkinase.			
DR	InterPro: IPR004019; YLP_motif.			
DR	Pfam: PF00757; Furin-like; 1.			
DR	Pfam: PF00069; pkinase; 1.			
DR	Pfam: PF01030; Recep_L_domain; 1.			
DR	Pfam: PF02757; YLP_2.			
DR	PRINTS: PR00109; TYRKINASE.			
DR	ProDom: PD000001; Euk_pkinase; 1.			
DR	SMART: SM00261; FU; 3.			
DR	SMART: SM00219; TYRK; 1.			
DR	PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.			
DR	PROSITE: PSS0011; PROTEIN_KINASE_DOM; 1.			
DR	PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.			
DR	PROSITE: PS00652; TNFR_NFR_1; UNKNOWN_1.			
KW	Kinase; Tyrosine-protein kinase.			
FT	NOM_TER	1		
SQ	SEQUENCE	1137 AA; 127927 MW; 4D616436F87DC84F CMC64;		
<hr/>				
Query Match		29.28;	Score 1156.5; DB 13; Length 1137;	
Best Local Similarity		42.6%;	Pred. No. 2.6e-95;	

Matches	207;	Conservative	80;	Mismatches	186;	Indels	7;	Gaps	6;
QY	161	LCYQDTILMKDIEFKNNQDALTLLIDFNSRACHPCSPMCKSGRCWGESSEDCQSILTRIVC	220						
Db	3	LCFADTTHMODIVRPMWMSNFTLVPTNGSSGCCGRCHKSCGTG-RCMGPTREHNCQTLTKTYC	61						
QY	221	AGGC-ARCKGRLPTCCCHGQCAAGCTGPRKHSCLACIHNHNSIGELHCPALVTYNTDIF	279						
Db	62	AEQCGRCYGPVPSQCHRECAGSGSPKDYDCFCACMNFNDSGACVYQCPQTFVYNPTTF	121						
QY	280	ESMPNDEGRYTCAGCAVTAAPYNYLSTVDGSCTLICPLINQDEVTAEADTORCKSCPCA	339						
Db	122	QLEHNHNAKYTGACVCKKCPHNFV-VBSSCVRACPSKMEV-BENGKMKCKPCTDICP	179						
QY	340	RVCYGLGMEHLREVAVTASNIQERAGCKKIFGSLAFIPESFDGDPASNTAPLQEPQLOV	399						
Db	180	KACDIDIGTGLSVSAQTVSSNIDKFINCTKINGNILEFLTGTGHGDPYHTIAAINPEKLN	239						
QY	400	FETTELITGYLISAMPDPLSLAEONQVIRGRILHGAVALSLTLOGISMLGRSR	459						
Db	240	FQTVETLITGYLISQWPEMMDERFVSNLVTIGGRALVSGLSLLTIKQGITSLQPSQSK	299						
QY	460	ELGSGTALIHNTHLCEFYTVPMWDLFRPHQALLHTANRPDECEVGEGLACHQLCARGH	519						
Db	300	QISAGNIITTDNSNLCYHTHTVMTSLFSPQSKTYIHRKKAKENCTADQMCNELCSSG	359						
QY	520	CWGPPTQCVNCSQPLRQGEVECEGRVLQGLREXYVNAHRCILPCHECOP-ONGSVTCFG	578						
Db	360	CWGPPTQCVNCSQPLRQGEVECEGRVLQGLREXYVNAHRCILPCHECOP-ONGSVTCFG	419						
QY	579	PEADQVACAHKKDPPEVCYARCPSEGVKRPDLSTMPYWKPFPEDEGACQPCINCTHSC	634						
Db	420	PGPDHCTKCFHFKDPNCPNCEKCPDGLQAGNSF-IFKVADEDRCHPCHPNCTQGC	473						
RESULT 10									
ID	Q90836	PRELIMINARY;	PRT;	527	AA.				
AC	Q90836;								
DT	01-NOV-1996 (Tremblrel. 01, Created)								
DT	01-NOV-1996 (Tremblrel. 01, Last sequence update)								
DT	01-JUN-2002 (Tremblrel. 21, Last annotation update)								
DE	EGF/TFG-alpha receptor precursor.								
GN	C-RRB.								
OS	Gallus gallus (Chicken).								
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;								
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;								
OC	Gallus.								
OX	NCBI_TaxID=9031;								
RN	[1]								
RP	SEQUENCE FROM N.A.								
RX	MEDLINE=92123214; PubMed=1732751;								
RA	Flickinger T.W., Maibhe N.J., Kung H.-J.;								
RT	"An alternatively processed mRNA from the avian c-erbB gene encodes a								
RT	soluble, truncated form of the receptor that can block ligand-								
RT	dependent transformation."								
RL	Mol. Cell. Biol. 12:883-893(1992).								
DR	EMBL: M77637; AAA48759.1; -								
DR	InterPro: IPR000494; EGFR_L_domain.								
DR	InterPro: IPR002174; Furin-like.								
DR	Pfam: PF00757; Furin-like; 1.								
DR	Pfam: PF01030; Recep_L_domain; 2.								
DR	SMART: SM00261; FU; 2.								
KW	Receptor; Signal.								
FT	SIGNAL	1	28	POTENTIAL.					
FT	CHAIN	29	527	EGF/TFG-ALPHA RECEPTOR.					
FT	SEQUENCE	527	AA;	58353	MM;	764564	AABC095298	CRC64;	
QY	Query Match	29.2%;	Score 1155;	DB 13;	Length 527;				
	Best Local Similarity	44.0%;	Pred. No. 1.3e-95;						
	Matches 227;	Conservative	87;	Mismatches 182;	Indels 20;	Gaps			
	11	LLALLPAGAST----	OVCYGTGDMKRLRASPETHLIDMLRHLYQGCQVVGNGNELTYL	65					

```

Db 20 LLLLLGRAVLCASVEEKKVCGCTNNKLLQGLGHVEDHFTSLQRYNNCEVLSNLETTY 79
Qy 66 PPNASISFLDIOEOYGVYLIANOVROYPLQRLRYRGOLFEDYVYALAVLNGDPLNN 125
Db 80 EHRHDLTFLTKTQEVAGVYLIANMVDVLPLENLQIRGVYLVADNSFALAVLSNYH-MNK 138
Qy 126 TTPVYGASPGGLRELOLRSTTEILKGVLIQIRNPOLCYODTILMKDIFHKNNOLATLID 185
Db 139 TQ-----GLRELPMKRLSELINGVKISNNPKLCNMDTYAMNDIITDSRK-PLTVLD 189
Qy 186 -TNRSPACHPCSPMKSGKSGWSESSDQSLRTVCAGGA-RCKGPLPTDCHEQCAAG 243
Db 190 FASNLSSCKCHNCTREDHCWGEQNCOTLTFVICAQCCSGRCGRKVPSDCCNCCAG 249
Qy 244 CGPKHSDCLACHENHSGICELHCPALTYNNDTFESMPNPRGRTFGASCTACPYNN 303
Db 250 CTGPRSDCLACKKEDDATTCKDTCPLVLYNFTTYQMDVNPBGKYSFGATCVRECPHN 309
Qy 304 LSTDVGSCTLVCPRLHNOEYTAEDGTQRCERKSPCARVCYGLGMEHLREVRATVSANIOE 363
Db 310 VYTDHSGVSCVSCNTDYEV-EENGVRKCKKCDGLSKVCNGLIGLKLISINATNIDS 368
Qy 364 FAGCKIFESLAFPEPSFGDPASNTAPLOPELOVFTLEETGYLYISAMPDSLPLDS 423
Db 369 FKNCTKINDVSTLIPYAFGLDAFTKTLPLDPKLDVFRYVKEISGLFIQAMPDNTADLY 428
Qy 424 VFONLOVGRILHNGAYSLTLOGLSIGSLRELGSGLALHHNTHLCFVHVPMD 483
Db 429 APENLLEIIRGRKHOGYSLAVYNLKIQSLGLSLKEISDGLAIKNNKLYADTMNR 488
Qy 484 QLEFRNPQALLHTANRPEDECVGEGSLACHOLCARGH 519
Db 489 SLFATOSQKTIIQNNRKNDC--SKSVCPAFAKAH 522

RESULT 11
Q9BIH9 PRELIMINARY: PRT; 1433 AA.
AC 09BIH9;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, last sequence update)
DE 01-JUN-2002 (Tremblrel. 21, last annotation update)
DE Putative epidermal growth factor receptor (Fragment).
GN EGFR.
OS Anopheles gambiae (African malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
OC Anopheles.
OX NCBI_taxid=7165;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-SUA;
RA Lycett G.J.;
RT "Cloning, expression and localisation of the Anopheles gambiae
RT epidermal growth factor receptor."
RT Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ301655; CAC35008.1; -.
DR HSSP; P1362; 1FGF.
DR InterPro; IPR000345; Cytc_heme_bind.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR Prodom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 7.
DR SMART; SM00220; S_TKc; 1.
DR SMART; SM00219; TYRKc; 1.

```

```

DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN 4.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00011; PROTEIN KINASE DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Receptor.
FT NON_TER
SQ SEQUENCE 1 159585 MW; E3D9D88967724F07 CRC64;

Query Match
Best Local Similarity 33.8%; Pred. No. 8; 3e-84; Length 1433;
Matches 233; Conservative 112; Mismatches 258; Indels 87; Gaps 21;

Qy 26 CTGTDKRLRLPASPEYTHLMRLHYOGCOVVOGNGLELYLPNASTSFLDIOEOYGVY 85
Db 1 CIGTNGRMGVNANREHYNNKRLNDRYINCYYVGNLEITMIONLTNDLNFQIHLREYGYL 60
Qy 86 IAHNOVROYPLQRLRYRGOLF-----EDNTALAVLNGDPLNNTPTVYGASPGGLREL 140
Db 61 ISLYDLPOVILPRLQIIRGRITTFKLNKWEAVGLFV-----SFSHMNTL 104
Qy 141 QLRLTEILKGGVLIQIRNPOLCYODTILMKDI-FHKNNOLATLIDTNRSPACHPCSPMC 199
Db 105 ELPALNDIIGSVGFNNYNLCHKMSINWBEILLAPQTSKOTTFNFSPEFYVCPCHPSC 164
Qy 200 KGSRCWSESSDQSLRTVCAGGA-RCKGPLPTDCHEQCAAGCTGPKHSDCLACH 257
Db 165 EVG-CWGECAHNQVRSKLNCSPOSGRCFGKPRECCHLFCAGGCTGPTGSDCLACKN 223
Qy 258 FHNSGICELHCPALTYNNDTFESMPNPRGRTFGASCTACPYNLTSDVGSCTLVCP 317
Db 224 FYDDGVCKQECPRMOLYNPTNFWEPNPDGKAYGATVYRKCP-EHLKDNACVYKCPK 282
Qy 318 HNOEYTAEDGTQRCERKSPCARVCYGLGMEHLREVRATVSANIOEFAGCKIFESLAF 377
Db 283 GKMPQNSE-----CVPCKGVCPKTCGEGIVH-----SDNIGYKXCTIIEGSELT 329
Qy 378 PESFGDPPASNT-----APLOPELOVFTLEETGYLYISAMPDSLPLDSVFNQIQ 429
Db 330 DQSFDFQOVYVNFSPGPRYIKIDPDLREVFSTVKEITGFINQAHHPFTYLNFRNLE 389
Qy 430 VIRGRILHNGAY-SLTLQGLSIGSLRELGSGLALHHNTHLCFVHVPMDLFPN 488
Db 390 VVGGRQLENLRASVITYVTSLSKLSLKRNSGSIVYLENSDLCFVEDIDWSEIKKS 449
Qy 489 PHQALLHTANRPEDECVGEGSLACHOLCARGHGWGPPTCCVNCQSLRQGEVEEGRVQ 548
Db 450 SBEVWVQNNRNATTEHEGMECSQKAGCMGKREPCLECKNNKYKGCIDSK--- 506
Qy 549 GLPREY-VNARHCLPCHPEQDPONGSVTCGPEADCCVACAHKDPFVCVAPSPGVKD 607
Db 507 SLPRLYSVDSKTCGDCHQCKD----FCYGNEDNCSCMVVKDGRFCVACDPT-TRHA 560
Qy 608 LSYMPTKRPDEGACQPCPINCTHSCVDLDKGC--PAQRASP--LTSQEDGDGPASP 663
Db 561 MN-----GTC---INCHKTCV-----GCKGPRDTJLAPDGCISCDKALIGS--- 597
Qy 664 LDSTFYSLLEDMDMDGLDAEYLVPOOG 693
Db 598 -DAKIERCLMKDSCPDGY-YSDYVLAQEEG 625

RESULT 12
Q9SESO PRELIMINARY: PRT; 478 AA.
AC 09SESO;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-OCT-2001 (Tremblrel. 18, last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, last annotation update)
DE Epidermal growth factor receptor related protein.
GN ERBP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

```



```

OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=GASTRO-DUODENAL MUCOUS;
RA Yu Y., Moshier J.A., Majumdar A.P.N.;
RT "Cloning of a novel Egfr-related peptide: A putative negative
   regulator of Egfr.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF187818; AAG17037.2; -
DR InterPro: IPR000494; EGFR_L_domain.
DR InterPro: IPR002174; Furin-like.
DR Pfam: PF00757; Furin-like; 1.
DR Pfam: PF01030; Recep_L_domain; 2.
KW Receptor.
SQ
SEQUENCE 478 AA; 53233 MW; CF873A8376C519E5 CRC64;

Query Match      26.0%; Score 1028.5; DB 11; Length 478;
Best Local Similarity 42.4%; Pred. No. 3,1e-84;
Matches 208; Conservative 69; Mismatches 167; Indels 47; Gaps 9;

OY 11 LLLALLPPGAA--STQVCTGDMKRLRLPASPETHLDMRLHLYGCGOVQGNLELYLPN 68
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 14 LITLALCAAGALEKKVQGSNRLTQLGTEDHFLSLQRMYNCEVVLNLEITYVGRN 73
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

OY 69 ASLSFLQDIOEVGYVLLAHNOVROVPLQRLRIYRGTOLEFEDNYALAVLNDGDPNNTTP 128
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 74 YDLSLTKIOEVAGFLIALTMTVERIPSEDQIIRGNALYENTYVALISN----- 124
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

OY 129 VTGASPGGLRELQLSLTEILKGVLLIQRPOLCYQDTILMKDI---FHKNNQALTLI 184
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 125 -YGNRTGLRELPHRNLOEILIGAVRFNNPILCNMDTIQWRDIQVNFMSNMSMDL--- 180
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

OY 185 DTNSRACHPSPCKGRCGSESEDCQSLTRYVCAAGCA-RCKGRLPTDCHQCCAG 243
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 181 -QSHSPCKDPSPCKGRCGSESEDCQSLTRYVCAAGCA-RCKGRLPTDCHQCCAG 239
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

OY 244 CTGPKHSQCLALPHNHSIGELHCPALVTYNTDTFESMPNPEGRYTGASCVTACPNY 303
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 240 CTGPKHSQCLALPHNHSIGELHCPALVTYNTDTFESMPNPEGRYTGASCVTACPNY 299
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

OY 304 LSTDVGSCTLVCPRLHNOEVTAEQTCQCEKSKPCARVCYGLGMEHLREAVTASANTOE 363
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 300 VYTDHSGCVRAKGPYEVV-EEDGIRKCKKCDGPKRKYCNNGIGEFKDTLSINATNIK 358
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

OY 364 FAGCKKIGSLAFLESPDGDPSANTAPLOEQLOVFELEITGYLYISAMPDLSPLDS 423
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 359 FRCTAISGDHLIPVAKGDSFTPTPLDPRELEIKETVETGSLLIQAMPENWTDLH 418
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

OY 424 VFONLQVIRGRILHNGAYSLTLQGLGI-----SMLGL-RSLRELGSGLALIHNT 472
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 419 AFENLEIRGRTRKQKQPSLAVVGLNITSLPWQVPSLSMQAVTRPLHPLAQ----- 469
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

OY 473 HLCFVHTVPM 483
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 470 -----NRVSWD 475
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 13
O9PSH2 PRELIMINARY; PRT; 599 AA.
AC O9PSH2;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Epidermal growth factor receptor (Fragment).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94020816; PubMed=8414496;

```

```

RA Callaghan T., Antczak M., Flickinger T., Raines M., Myers M.,
RA Kung H.J.;
RT "A complete description of the Egfr-receptor exon structure:
   implication in oncogenic activation and domain evolution.";
RL Oncogene 8:2939-2948(1993).
DR InterPro: IPR000494; EGFR_L_domain.
DR InterPro: IPR002174; Furin-like.
DR Pfam: PF00757; Furin-like; 2.
DR Pfam: PF01030; Recep_L_domain; 2.
DR SMART: SM00261; FU_3.
SQ
SEQUENCE 599 AA; 66363 MW; FEAB46D293D991BD CRC64;

Query Match      24.6%; Score 971.5; DB 13; Length 599;
Best Local Similarity 30.2%; Pred. No. 5,8e-79;
Matches 222; Conservative 76; Mismatches 180; Indels 257; Gaps 11;

OY 24 QVCTGDMKRLRLPASPETHLDMRLHLYGCGOVQGNLELYLPNASLSFLQDIOEVGY 83
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5 KYCQGTNNKLTQLGHVEHFTSLQRMYNCEVVLNLEITYVEHNRDLFLKTOIEVAGY 64
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

OY 84 VLIANQROYVPLQRLRIYRGTOLEFEDNYALAVLNDGDPNNTTPVTGASPGGLRELQLR 143
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 65 VLIANMVDVPLLENLQIRGNVLYDNSEFALAVLSNYH-MNKTQ-----GLRELPMK 115
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

OY 144 SLTEILKGVLLQRPOLCYQDTILMKDIFHKNNQALTLIDTNSRACHPSPCKGSR 203
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 116 RLSEILNGSVKISNNPKLCNMDTVLMDI-----IDTSK----- 150
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

OY 204 CWESSEDCQSLTRYVCAAGCARCKGRLPTDCHQCAAGCTGPKHSQCLALPHNHSIGI 263
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 151 -----PL----- 152
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

OY 264 CELHCPALVTYNTDTFESMPNPEGRYTGASCVTACPNYLSSTDVGSCTLVCPRLHNOEVT 323
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 153 -----TVDFASNLSSV 164
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

OY 324 AEDGTQRCCKSPCARVCYGLGMEHLREAVTASANTOEFAGCKKIGSLAFLESPDG 383
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 165 EENGVRKCKKCDGLSKCKNGIGELGILSINATNIDSRKNCIKNGVSLIPVALLG 224
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

OY 384 DPASNTAPLOEQLOVFELEITGYLYISAMPDLSPLSVFONLQVIRGRILHNGAYSL 443
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 225 DAFTYTLPLDPKLDVFTVKEISGFLLIQAMPDN-ATLYAFENLEIRGRTRKQKQYSL 283
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

OY 444 TLQGLIGTMLGRSLRELGSGLALIHNTHLCFVHTVPMQDLFRPHALHTAR----- 499
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 284 AVVNLKIQSLGRSLKEISDGDIALMKNNLCYADTMNRSFLFATQSQTKIIONRRKND 343
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

OY 500 -----PEDECVGEL-----ACHQLCARG----- 518
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 344 CCRKHPNCTEDHCGAGBQNKQTLTKYICAQCCGRCGRKGVPSDCCINQCAAGCTGPRE 403
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

OY 519 ----- 518
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 404 SDGLACRFRDDATCKDTCPLVLYXNPTTYQMDVNPBEKYSFGATCVRECPHNTADRHVC 463
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

OY 519 -----HNGRGPRTQVCNCSQFLRGQCEYECRYLQGLPREYVNAKICLCHPCQON 571
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 464 DPLGSDVGCWGGPRPHCSRFKQCEVCVQCNILLQGEPRFEEDSKCLCHSSECLQON 523
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

OY 572 G----SVTCFGEADQCVCAHAKYDPPFCVARGSPGVKFDLYMPIMKPPDEGACQPCPI 628
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 524 STAYVTTSSGGRPDHGMCAHIFIDGPRHCYKACPAVAGENDTL-VKMTADANNAQCQCHP 582
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

OY 629 NCTHSCVDLDDKCP 643
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 583 NCTRCKKGPGLGCP 597
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 14
O923V5 PRELIMINARY; PRT; 176 AA.
AC O923V5;
RX MEDLINE=94020816; PubMed=8414496;

```

DT 01-DEC-2001 (TREMblrel. 19, Created)  
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE Epidermal growth factor receptor (Fragment).  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 NCBI\_TaxId=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MISTAR;  
 RA Venzani F.M., Soverchia L., Concetti A.;  
 RT "Rat HER-2/neu oncogene sequence: unfaithful nucleotide assignment."  
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF393158; AAK76996.1;  
 KM Receptor.  
 FT NON\_TER 1 1  
 FT NON\_TER 176 176  
 SQ SEQUENCE 176 AA; 19491 MW; 80C6594C3EBE32A CRC64;

Query Match 22.4%; Score 887; DB 11; Length 176;  
 Best Local Similarity 83.5%; Pred. No. 5.1e-72;  
 Matches 147; Conservative 11; Mismatches 18; Indels 0; Gaps 0;

QY 477 VHTVPMQDLFRNPHQALLHTANRPEDCVGEGLCACHQLCARGHCWGPGPTQCVNCSQFLR 536  
 DB 1 VHTVPMQDLFRNPHQALLHTANRPEDCVGEGLCACHQLCARGHCWGPGPTQCVNCSHFLR 60  
 QY 537 GQECVCECRVLOGIPREYVNAHRLPCHECOPONGSVTCGCPREADOCVACAHYKDDPEPC 596  
 DB 61 GQECVCECRVWKGLPREVSVSKRCLPCHECOPONSSETCGSEADQCAACAHYKDDSSSC 120  
 QY 597 VARPSCVKKPDLSTYMPIMKFPDEEGACOPCPINCTHSCVDLDDKCPAPQASPLT 652  
 DB 121 VARPSCVKKPDLSTYMPIMKFPDEEGICQPCPINCTHSCVDLDERGCAEQASAPVT 176

RESULT 15  
 Q9BUD7  
 ID Q9BUD7 PRELIMINARY; PRT; 331 AA.  
 AC Q9BUD7;  
 DT 01-JUN-2001 (TREMblrel. 17, Created)  
 DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)  
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)  
 DE Similar to v-erb-b2 avian erythroblastic leukemia viral oncogene homolog 3.  
 OS Homo sapiens (human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NCBI\_TaxId=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=PLACENTA;  
 RA Strausberg R.;  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC002706; AA02706.1;  
 DR InterPro; IPR000494; EGFR\_L\_domain.  
 DR InterPro; IPR002174; Furin-like.  
 DR Pfam; PF00757; Furin-like.1.  
 DR Pfam; PF01030; Recep\_L\_domain.1.  
 DR SMART; SM00261; FU; 2.  
 SQ SEQUENCE 331 AA; 36489 MW; 45B8BEBE683FE7E8 CRC64;

Query Match 18.6%; Score 734; DB 4; Length 331;  
 Best Local Similarity 44.0%; Pred. No. 7.5e-58;  
 Matches 147; Conservative 49; Mismatches 120; Indels 18; Gaps 8;

QY 10 GLIALALPPGAA--STVCTGTDMKRLRPASPETHLMLRLHYGCGVOCNLELYLPT 67  
 DB 11 GLPFLARGSEVNSQAVCEGTGLNGLSVTGDAENQYOTLYKLYRCEVWGNLEIYLTGH 70  
 QY 68 NASLSFLQDIOEVQGYVLIANNOVYPLQRLRIVRGTQLEEDNYALAVIDNGDPLNNTT 127

DB 71 NADISFLQWIREVTGYVLVAMNEFSTLPDNLRYVRGTQYDGFALFV-----LNYNT 125  
 QY 128 PVTGASPGGLRELOLRSLTEILKGVLIQRPOLCYODTILMKDIFKKNQALATLIDTN 187  
 DB 126 ----NSSHALRQLRLTQLTETLSGVYIERKNDKCHMDITDMRDIVARDR---AEIYKD 178  
 QY 188 RSRACHPCSPCKGSRGWGESSDQSLFTVTCAGG-ARCKGPLPTDCCHEQCAAGCTG 246  
 DB 179 NGRSCPCHVEYCKG-RCWGPSEDCQTLTKTICAPQCNGHCFCGPNQCCCHDECAGGCSG 237  
 QY 247 PKHSDCLACLHFNNSGICELHCPALVYNTDTPESMPNPEGRYTFGASCVTACPYNTLS 306  
 DB 238 PQDTDFCFCRHFNDGACVPCPQPLVYKLTQLEPNPHTKYQYGGVAVASCPHNFV-V 296  
 QY 307 DVGSCITLVCPLHNOEYTAEDGTQRCCKSKPCAR 340  
 DB 297 DQTSVCACPCPDKKMEVD-KNGLKMCCECGGLCPK 329

Search completed: January 13, 2003, 14:48:29  
 Job time : 36.7407 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 13, 2003, 14:44:10 ; Search time 12.3961 Seconds

(without alignments)  
3074.904 Million cell updates/sec

Title: US-09-854-356-6

Perfect score: 5078

Sequence: 1 MELALCRKGLLALLPPGA.....TFKGTPTAENPEYGLDVPV 919

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40\*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4900	96.5	1255	1	ERB2_HUMAN
2	4134	81.4	1257	1	ERB2_RAT
3	4122.5	81.2	1254	1	ERB2_MESAU
4	1643	32.4	1210	1	EGFR_HUMAN
5	1612	31.7	1210	1	EGFR_MOUSE
6	1528.5	30.1	1308	1	ERB4_HUMAN
7	1519	29.9	1308	1	ERB4_RAT
8	1509.5	29.7	703	1	EGFR_CHICK
9	1435	28.3	1342	1	ERB3_HUMAN
10	1374.5	27.1	1339	1	ERB3_RAT
11	1330	26.0	1167	1	XMRK_XIPMA
12	1013.5	20.2	1426	1	EGFR_DROME
13	718.5	14.1	1323	1	LT23_CAEL
14	427	8.4	245	1	ERB2_MOUSE
15	390.5	7.7	1300	1	IRR_MOUSE
16	384	7.6	1363	1	ILPR_BRAIA
17	377.5	7.4	1300	1	IRR_CAVPO
18	368	7.2	1382	1	INSR_HUMAN
19	364	7.2	581	1	IRR_RAT
20	362.5	7.1	1383	1	INSR_RAT
21	361	7.1	1372	1	INSR_MOUSE
22	359	7.1	1297	1	IRR_HUMAN
23	353	7.0	1477	1	HHR7_HYDAT
24	326	6.4	1390	1	INSR_AEDAE
25	324	6.4	1607	1	MLPR_LYMET
26	316	6.2	1370	1	IGLR_RAT
27	314	6.2	1373	1	IGLR_MOUSE
28	311.5	6.1	1367	1	IGIR_HUMAN
29	307	6.0	2146	1	INSR_DROME
30	288	5.7	634	1	ERBB_ALV
31	271.5	5.3	1696	1	PKCS_BRACL
32	254.5	5.0	604	1	ERBB_AVIER
33	244.5	4.8	1877	1	PKCS_MOUSE

34	208	4.1	1680	1	FUR2_DROME	P30432 dtrosophila
35	196	3.9	5179	1	MUC2_HUMAN	Q02817 homo sapien
36	195.5	3.8	937	1	PAC4_RAT	Q63415 rattus norv
37	195	3.8	830	1	SREC_HUMAN	Q14162 homo sapien
38	186.5	3.7	540	1	ERBB_AVIER	P11273 avian eryth
39	184.5	3.6	913	1	PKCS_HUMAN	Q92824 homo sapien
40	180.5	3.6	1877	1	PKC5_RAT	P41413 rattus norv
41	176	3.5	969	1	PAC4_HUMAN	P29122 homo sapien
42	175.5	3.5	2003	1	NRC4_HUMAN	Q99466 homo sapien
43	172.5	3.4	1173	1	TSPI_XENLA	P35448 xenopus lae
44	170	3.3	3312	1	CUR3_HUMAN	Q9HY97 homo sapien
45	170	3.3	3707	1	PGSM_MOUSE	Q05793 mus musculu

## ALIGNMENTS

RESULT 1  
ID ERB2\_HUMAN STANDARD: PRT: 1255 AA.  
AC P04626;  
DT 13-AUG-1987 (Rel. 05, Created)  
DT 13-AUG-1987 (Rel. 05, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)  
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (tyrosine kinase-type cell  
DE surface receptor HER2) (MLN 19).  
GN ERB2 OR HER2 OR NGL OR NEU.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_Taxid:9606;  
[1]  
RP MEDLINE=86118663; PubMed=3003577;  
RA Yamamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N.,  
RA Saito T., Toyoshima K.,  
RT "Similarity of protein encoded by the human c-erb-B-2 gene to  
RT epidermal growth factor receptor.";  
RT Nature 319:230-234(1986).  
[2]  
RN SEQUENCE FROM N.A.  
RP MEDLINE=86070181; PubMed=2999974;  
RX Coussens L., Yang-Feng T.L., Liao Y.C., Chen E., Gray A.,  
RA McGath J., Seeburg P.H., Libermann T.A., Schlessinger J.,  
RA Francke U., Levinson A., Ullrich A.;  
RT "Tyrosine kinase receptor with extensive homology to EGF receptor  
RT shares chromosomal location with neu oncogene.";  
RT Science 230:1132-1139(1985).  
[3]  
RN SEQUENCE OF 737-1031 FROM N.A.  
RP MEDLINE=86016729; PubMed=2995967;  
RX Semba K., Kamata N., Toyoshima K., Yamamoto T.;  
RT "A v-erbB-related protooncogene, c-erbB-2, is distinct from the  
RT c-erbB-1/epidermal growth factor-receptor gene and is amplified in a  
RT human salivary gland adenocarcinoma.";  
RT Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501(1985).  
[4]  
RX VARIANTS VAL-654 AND VAL-655.  
RP MEDLINE=93194196; PubMed=8095488;  
RA Ehsani A., Low J., Wallace R.B., Wu A.M.;  
RT "Characterization of a new allele of the human ERBB2 gene by allele-  
RT specific competition hybridization.";  
RL Genomics 15:426-429(1993).  
CC -1- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,  
CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A  
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-  
CC ALPHA AND AMPHIREGULIN.  
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
CC tyrosine phosphate.  
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS  
CC (POTENTIAL).  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.

CC -1- P-TM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE  
CC RESIDUES (BY SIMILARITY).  
CC -1- POLYMORPHISM: THERE ARE FOUR ALLELES DUE TO THE VARIATIONS IN  
CC POSITIONS 654 AND 655. ALLELE B1 (654-ILE-ILE-655) HAS A FREQUENCY  
CC OF 0.783; ALLELE B2 (654-ILE-VAL-655) HAS A FREQUENCY OF 0.206;  
CC ALLELE B3 (654-VAL-VAL-655) HAS A FREQUENCY OF 0.012.  
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
CC -----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: M11767; AAA35808.1; -.  
DR EMBL: M11761; AAA35808.1; JOINED.  
DR EMBL: M11762; AAA35808.1; JOINED.  
DR EMBL: M11763; AAA35808.1; JOINED.  
DR EMBL: M11764; AAA35808.1; JOINED.  
DR EMBL: M11765; AAA35808.1; JOINED.  
DR EMBL: M11766; AAA35808.1; JOINED.  
DR EMBL: M11730; AAA75493.1; -.  
DR EMBL: M12036; AAA35978.1; -.  
DR EMBL: X03363; CAA27060.1; -.  
DR PIR: A25491; A25491.  
DR PIR: A24571; A24571.  
DR HSP: P11362; 1FGK.  
DR GeneW: HGNC:3430; ERBB2.  
DR MIM: 164870; -.  
DR InterPro: IPR000494; EGFR\_L\_domain.  
DR InterPro: IPR000719; Euk\_pkinase.  
DR InterPro: IPR001245; Tyr\_pkinase.  
DR InterPro: IPR004019; YLP\_motif.  
DR Pfam: PF00069; pkinase.1.  
DR Pfam: PF00757; Furin-like.1.  
DR Pfam: PF01030; Recep\_L\_domain.2.  
DR Pfam: PF02757; YLP.2.  
DR ProDom: PD000001; Euk\_pkinase.1.  
DR SMART: SM00219; Fv; 3.  
DR SMART: SM00219; TyrKc.1.  
DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP.1.  
DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR.1.  
DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM.1.  
KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;  
KW Transference; Tyrosine-protein Kinase; ATP-binding; Phosphorylation;  
KW Polymorphism.  
KM SIGNAL. 1 21  
FT CHAIN 22 1235 POTENTIAL. RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.  
FT DOMAIN 22 652 EXTRACELLULAR (POTENTIAL).  
FT TRANSLEM 653 675 POTENTIAL.  
FT DOMAIN 676 1255 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 720 987 PROTEIN KINASE.  
FT NP\_BIND 726 734 ATP (BY SIMILARITY).  
FT BINDING 733 733 ATP (BY SIMILARITY).  
FT ACT\_SITE 845 845 BY SIMILARITY.  
FT DISULFID 195 204 BY SIMILARITY.  
FT DISULFID 199 212 BY SIMILARITY.  
FT DISULFID 220 227 BY SIMILARITY.  
FT DISULFID 224 235 BY SIMILARITY.  
FT DISULFID 236 244 BY SIMILARITY.  
FT DISULFID 240 252 BY SIMILARITY.  
FT DISULFID 255 264 BY SIMILARITY.  
FT DISULFID 268 295 BY SIMILARITY.  
FT DISULFID 299 311 BY SIMILARITY.  
FT DISULFID 315 331 BY SIMILARITY.  
FT DISULFID 334 338 BY SIMILARITY.  
FT DISULFID 511 520 BY SIMILARITY.  
FT DISULFID 515 528 BY SIMILARITY.  
FT DISULFID 531 540 BY SIMILARITY.  
FT DISULFID 544 560 BY SIMILARITY.

FT DISULFID 563 576 BY SIMILARITY.  
FT DISULFID 567 584 BY SIMILARITY.  
FT DISULFID 587 596 BY SIMILARITY.  
FT DISULFID 600 623 BY SIMILARITY.  
FT DISULFID 626 634 BY SIMILARITY.  
FT DISULFID 630 642 BY SIMILARITY.  
FT MOD\_RES 1139 1139 BY SIMILARITY.  
FT CARBOHYD 1248 1248 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
FT CARBOHYD 124 68 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 187 187 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 239 259 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 530 530 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 571 571 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 629 629 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT VARIANT 654 654 I -> V.  
FT VARIANT 655 655 /FTID-VAR\_004077.  
FT VARIANT 655 655 I -> V.  
FT CONFLICT 1170 1170 /FTID-VAR\_004078.  
FT SEQUENCE 1255 AA; 137909 MW; 39E9DEDA04DCP962 CRC64;  
SQ  
Query Match 96.5%; Score 4900; DB 1; Length 1255;  
Best Local Similarity 73.2%; Pred. No. 1.2e-275;  
Matches 919; Conservative 0; Mismatches 0; Indels 336; Gaps 1;  
QY 1 MELALCRWGLLALLPPGAASVCTGTDKRLRPSPTHLDMLRHLXGCGVVOGNTL 60  
DB 1 MELALCRWGLLALLPPGAASVCTGTDKRLRPSPTHLDMLRHLXGCGVVOGNTL 60  
QY 61 ELTYLPINASISFLQDIOEVQGYVLIANVOYRVPDLRLRVRGTOLEEDNYALAVDNG 120  
DB 61 ELTYLPINASISFLQDIOEVQGYVLIANVOYRVPDLRLRVRGTOLEEDNYALAVDNG 120  
QY 121 DPLNNTPTVTCASPGGLRELDRLSTELTKGVIQIRNPOLCYDDTILMKDIFKNNOLA 180  
DB 121 DPLNNTPTVTCASPGGLRELDRLSTELTKGVIQIRNPOLCYDDTILMKDIFKNNOLA 180  
QY 121 DPLNNTPTVTCASPGGLRELDRLSTELTKGVIQIRNPOLCYDDTILMKDIFKNNOLA 180  
DB 121 DPLNNTPTVTCASPGGLRELDRLSTELTKGVIQIRNPOLCYDDTILMKDIFKNNOLA 180  
QY 181 LTLIDITNRSRACHPCSPWCKSGRCMGSSSEDCQSLRTVCAGGACRCKGRLPTDCHEQC 240  
DB 181 LTLIDITNRSRACHPCSPWCKSGRCMGSSSEDCQSLRTVCAGGACRCKGRLPTDCHEQC 240  
QY 181 LTLIDITNRSRACHPCSPWCKSGRCMGSSSEDCQSLRTVCAGGACRCKGRLPTDCHEQC 240  
DB 181 LTLIDITNRSRACHPCSPWCKSGRCMGSSSEDCQSLRTVCAGGACRCKGRLPTDCHEQC 240  
QY 241 AAGCTGKRHSCLACLFHNSGICELCPALVTYNTDTFESMPNBERYTFGASCVTACP 300  
DB 241 AAGCTGKRHSCLACLFHNSGICELCPALVTYNTDTFESMPNBERYTFGASCVTACP 300  
QY 241 AAGCTGKRHSCLACLFHNSGICELCPALVTYNTDTFESMPNBERYTFGASCVTACP 300  
DB 241 AAGCTGKRHSCLACLFHNSGICELCPALVTYNTDTFESMPNBERYTFGASCVTACP 300  
QY 301 YNYSTVSGSCVLCPLHNOEVTAEQRCCKSCARCYGLGMEHLREVAVTYSAN 360  
DB 301 YNYSTVSGSCVLCPLHNOEVTAEQRCCKSCARCYGLGMEHLREVAVTYSAN 360  
QY 301 YNYSTVSGSCVLCPLHNOEVTAEQRCCKSCARCYGLGMEHLREVAVTYSAN 360  
DB 301 YNYSTVSGSCVLCPLHNOEVTAEQRCCKSCARCYGLGMEHLREVAVTYSAN 360  
QY 361 IOEFAGCKRTFGSLAFPESEFDGDPASNTAPLQEOLOVFETLEITGYLISAMPDPLP 420  
DB 361 IOEFAGCKRTFGSLAFPESEFDGDPASNTAPLQEOLOVFETLEITGYLISAMPDPLP 420  
QY 361 IOEFAGCKRTFGSLAFPESEFDGDPASNTAPLQEOLOVFETLEITGYLISAMPDPLP 420  
DB 361 IOEFAGCKRTFGSLAFPESEFDGDPASNTAPLQEOLOVFETLEITGYLISAMPDPLP 420  
QY 421 DLSVFQNLQVIRGILINAGASLTLQGLISWGLRSLRELGSGLALIHNTHLCFVHTV 480  
DB 421 DLSVFQNLQVIRGILINAGASLTLQGLISWGLRSLRELGSGLALIHNTHLCFVHTV 480  
QY 421 DLSVFQNLQVIRGILINAGASLTLQGLISWGLRSLRELGSGLALIHNTHLCFVHTV 480  
DB 421 DLSVFQNLQVIRGILINAGASLTLQGLISWGLRSLRELGSGLALIHNTHLCFVHTV 480  
QY 481 PMDOLFNRPHQALHTANRPEDCEVGEGLACHQCLAGHCHGMPPTOCVNCOSQPLRQEC 540  
DB 481 PMDOLFNRPHQALHTANRPEDCEVGEGLACHQCLAGHCHGMPPTOCVNCOSQPLRQEC 540  
QY 481 PMDOLFNRPHQALHTANRPEDCEVGEGLACHQCLAGHCHGMPPTOCVNCOSQPLRQEC 540  
DB 481 PMDOLFNRPHQALHTANRPEDCEVGEGLACHQCLAGHCHGMPPTOCVNCOSQPLRQEC 540  
QY 541 VEECRVLQGLPREVYVNAHCLPCHECOPONGSVTCGPEADQCVACAHYKDPFCVARC 600  
DB 541 VEECRVLQGLPREVYVNAHCLPCHECOPONGSVTCGPEADQCVACAHYKDPFCVARC 600  
QY 541 VEECRVLQGLPREVYVNAHCLPCHECOPONGSVTCGPEADQCVACAHYKDPFCVARC 600  
DB 541 VEECRVLQGLPREVYVNAHCLPCHECOPONGSVTCGPEADQCVACAHYKDPFCVARC 600  
QY 601 PSQVKKPLSTYPIKPFDEBGAQPCPINCTHSCVDDDKGCPADQASPLTS----- 653  
DB 601 PSQVKKPLSTYPIKPFDEBGAQPCPINCTHSCVDDDKGCPADQASPLTS----- 653  
QY 601 PSQVKKPLSTYPIKPFDEBGAQPCPINCTHSCVDDDKGCPADQASPLTS----- 653  
DB 601 PSQVKKPLSTYPIKPFDEBGAQPCPINCTHSCVDDDKGCPADQASPLTS----- 653  
QY 654 ----- 653  
DB 661 ILLVYVLGVVFGILLKRRQOKIRKTYMRRLLOETELVEPLTPSGAMPNOAMRIKTEL 720

QY 654 ----- 653  
 DB 721 RKVKVLGSGAFSTYKGIWIPDGENVKIPAIKVLRENTSPKANKEILDEATVAVAGVSP 780  
 QY 654 ----- 653  
 DB 781 YVSRLLGICLTSTVOLYQLMPYGCILDHVRENRRGLSGQDLLNMCMQIAKMGSTLEQVR 840  
 QY 654 ----- 653  
 DB 841 LVHRDLAARNVLYKSPNHVKITDFGLARLLIDETEHADGKVPDKWMALESILRRPT 900  
 QY 654 ----- 653  
 DB 901 HOSDWSYGVTVWELMTFGAKPYDGIAPAREIPDLLEKGERLPDPICITDIYVIMVKKMM 960  
 QY 654 ----- 684  
 DB 961 IDSECRPRRELVSFESNRARDPQRFVYIQNEDLGPSALDSTFYRSLLDEDDMDGLDVA 1020  
 QY 685 EBYLVPOGFCPCDPAPAGAGVHHRRHSSSTRSGGGLTTLGLEPSEERAPRSPPLAPSEG 744  
 DB 1021 EBYLVPOGFCPCDPAPAGAGVHHRRHSSSTRSGGGLTTLGLEPSEERAPRSPPLAPSEG 1080  
 QY 745 AGSDVFDGDLGMAKGLQSLPTHPSPLOQRYSEDPVPLPSETDGYAAPLTCSPQPEYV 804  
 DB 1081 AGSDVFDGDLGMAKGLQSLPTHPSPLOQRYSEDPVPLPSETDGYAAPLTCSPQPEYV 1140  
 QY 805 NOPDVPRPSPREPRPLPAPAPAGATLERPKTLSPGKNGVADVAFAGAVENPELTPQ 864  
 DB 1141 NOPDVPRPSPREPRPLPAPAPAGATLERPKTLSPGKNGVADVAFAGAVENPELTPQ 1200  
 QY 865 GGAPQHPHPPEAFSPADNLVYWDODPPERCAPSPSTEGTPTAENPEYGLDVPV 919  
 DB 1201 GGAPQHPHPPEAFSPADNLVYWDODPPERCAPSPSTEGTPTAENPEYGLDVPV 1255  
 RESULT 2  
 ERBB2\_RAT  
 ID ERBB2\_RAT STANDARD: PRT: 1257 AA.  
 AC P06494;  
 DT 01-JAN-1988 (Rel. 06, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Receptor protein-tyrosine kinase erbb-2 precursor (EC 2.7.1.112)  
 DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Epidermal growth factor receptor-related protein).  
 DE Receptor-related protein).  
 GN ERBB2 OR NEU  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Neuroblastoma;  
 RX MEDLINE=86118662; PubMed=3945311;  
 RA Baigmann C.I., Hung M.-C., Weinberg R.A.;  
 RT "The neu oncogene encodes an epidermal growth factor receptor-related protein."  
 RT Nature 319:226-230(1986).  
 RL [2]  
 RP SEQUENCE OF 852-905 FROM N.A.  
 RC TISSUE-Sciatic nerve;  
 RX MEDLINE=9122560; PubMed=2025425;  
 RA Lai C., Lemke G.;  
 RT "An extended family of protein-tyrosine kinase genes differentially expressed in the vertebrate nervous system."  
 RT Neuron 6:691-704(1991).  
 RL [3]  
 RP STRUCTURE BY NMR OF 650-668.  
 RX MEDLINE=92155181; PubMed=1346763;  
 RA Gullick W.J., Bottomley A.C., Loftis F.J., Doak D.G., Mulvey D.,

RA Newman R., Crumpton M.J., Sternberg M.J.E., Campbell I.D.;  
 RT "Three dimensional structure of the transmembrane region of the proto-oncogenic and oncogenic forms of the neu protein."  
 RL EMBO J. 11:43-48(1992).  
 CC -I- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX, ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-ALPHA AND AMPHIREGULIN.  
 CC -I- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein tyrosine phosphate.  
 CC -I- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS. THE CONSTITUTIVELY ACTIVATED ONCOGENIC VARIANT FORMS A HOMODIMER.  
 CC -I- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -I- PMW: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES (BY SIMILARITY).  
 CC -I- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL, X03362; CA27059.1; ALF\_INIT.  
 DR PIR, A24562; TVRTNU.  
 DR HSSP, P11362; IEGK.  
 DR InterPro: IPR000494; EGFR\_L\_domain.  
 DR InterPro: IPR000719; Euk\_pkinase.  
 DR InterPro: IPR002174; Furin-like.  
 DR InterPro: IPR001245; Tyr\_pkinase.  
 DR InterPro: IPR004019; YLP\_motif.  
 DR Pfam: PF00069; pkinase; 1.  
 DR Pfam: PF00757; Furin-like; 1.  
 DR Pfam: PF01030; Recep\_L\_domain; 2.  
 DR Pfam: PF02757; YLP; 2.  
 DR ProDom: PD000001; Euk\_pkinase; 1.  
 DR SMART, SM00261; Fu; 3.  
 DR SMART, SM00219; Tyrc; 1.  
 DR PROSITE, PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE, PS00109; PROTEIN\_KINASE\_TYR; 1.  
 DR PROSITE, PS50011; PROTEIN\_KINASE\_DOM; 1.  
 KM Transmembrane; Glycoprotein; Multigene family; Receptor; Signal; Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation; Proto-oncogene; Disease mutation.  
 KM SIGNAL  
 FT CHAIN 1 21  
 FT DOMAIN 22 1257 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.  
 FT TRANSMEM 22 654 EXTRACELLULAR (POTENTIAL).  
 FT DOMAIN 655 677 POTENTIAL.  
 FT DOMAIN 678 1257 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 159 369 CYS-RICH.  
 FT DOMAIN 473 646 CYS-RICH.  
 FT DOMAIN 722 989 PROTEIN KINASE.  
 FT NP\_BIND 728 736 ATP (BY SIMILARITY).  
 FT BINDING 755 755 ATP (BY SIMILARITY).  
 FT ACT\_SITE 847 847 BY SIMILARITY.  
 FT DISULFID 196 205 BY SIMILARITY.  
 FT DISULFID 200 213 BY SIMILARITY.  
 FT DISULFID 221 228 BY SIMILARITY.  
 FT DISULFID 225 236 BY SIMILARITY.  
 FT DISULFID 237 245 BY SIMILARITY.  
 FT DISULFID 241 253 BY SIMILARITY.  
 FT DISULFID 256 265 BY SIMILARITY.  
 FT DISULFID 269 296 BY SIMILARITY.  
 FT DISULFID 300 312 BY SIMILARITY.  
 FT DISULFID 316 332 BY SIMILARITY.  
 FT DISULFID 335 339 BY SIMILARITY.  
 FT DISULFID 513 522 BY SIMILARITY.  
 FT DISULFID 517 530 BY SIMILARITY.  
 FT DISULFID 533 542 BY SIMILARITY.  
 FT DISULFID 546 562 BY SIMILARITY.  
 FT DISULFID 565 578 BY SIMILARITY.

FT	DISULFID	569	586	BY SIMILARITY.
FT	DISULFID	589	598	BY SIMILARITY.
FT	DISULFID	602	625	BY SIMILARITY.
FT	DISULFID	628	636	BY SIMILARITY.
FT	DISULFID	632	644	BY SIMILARITY.
FT	MOD. RES	1141	1141	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	MOD. RES	1250	1250	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	CARBOHYD	68	68	N-LINKED (GLCNAC . . .) (POTENTIAL).
FT	CARBOHYD	188	188	N-LINKED (GLCNAC . . .) (POTENTIAL).
FT	CARBOHYD	260	260	N-LINKED (GLCNAC . . .) (POTENTIAL).
FT	CARBOHYD	532	532	N-LINKED (GLCNAC . . .) (POTENTIAL).
FT	CARBOHYD	573	573	N-LINKED (GLCNAC . . .) (POTENTIAL).
FT	CARBOHYD	631	631	N-LINKED (GLCNAC . . .) (POTENTIAL).
FT	VARIANT	661	661	V -> E (IN ONCOGENIC NEU).
SQ	SEQUENCE	1257 AA;	138831 MM;	6129264583011402 CRC64;
Query Match		81.4%;	Score 4134;	DB 1; Length 1257;
Best Local Similarity		62.1%;	Pred. No. 2e-231;	
Matches 781;		Conservative 42;	Mismatches 96;	Indels 338; Gaps 3;
OY	1 MELALACRMGILLALLPRPAASTGYCTGDMTLRLPASPEHLMIMLRHIYOCQVVGNTL	60		
Dd	1 MELALMCRMGFLALLPRGIAGTGYCTGDMTLRLPASPEHLMIMLRHIYOCQVVGNTL	60		
OY	61 ELTYLPYNASLSFLDDIOGVGVYLIAHQVQVPLQRIRIYRGOLFEDNALAVLDNG	120		
Dd	61 ELTYLPANASLSFLDDIOGVGYMLIANHQVRVPLQRIRIYRGOLFEDKAKALAVLDNR	120		
OY	121 DPLNTPPYT-GASPGRELEQLRSLTEILKGVLIGNRPOLCYDTILMKDIFFKNNOI	179		
Dd	121 DDVVAASTPGPTREGELRELQRLSRITETILKGVLIRGNPOLCYDMVLMDKVFRNNOL	180		
OY	*180 ALLTLDTNSRACHCSPMKCSRCMGESSEDCQSILTRIVCAGGARCKGRPLPTCCCHQ	239		
Dd	181 AAYVDITDNSRACPAPAPACKDNHCWGESPCCQILTGTICTSGCARCKGRPLPTDCCHQ	240		
OY	240 CAAGCTGPKHSOCLACLFHNHSGICEHLCPALVTYTNTDFEEMPNPEGRTFGASCATAC	299		
Dd	241 CAAAGCTGPKHSOCLACLFHNHSGICEHLCPALVTYTNTDFEEMNPNEGRTTGASCATTGC	300		
OY	300 PYNVLSTDVGSCTLVCPRLNHQVEATADGTQORCEKSKPCARVCYGLGMHLEBVRAYTSA	359		
Dd	301 PYNVLTSTEVSCTLVCPRNNOEVTAEDGTQORCEKSKPCARVCYGLGMHLEBVRAYSISD	360		
OY	360 NIQERAGCKKTIGSLAFLEPSFDGPASTAARLPQPOLOVFETLEBTGYLIISAMPDL	419		
Dd	361 NVOEPDGCKKIGSLAFLEPSFDGPSISAIARLPQLOLVFEETLEBTGYLIISAMPDL	420		
OY	420 POLSVPONLOVIRGRILIHNGAYSLTLOGIGIMWLGIRSLRELGSGIALIHHHTHLCFVHT	479		
Dd	421 ROLSVPONLRIRGRILHDGAYSLTLOGIGISTLRSLRELGSGIALIHRHNAHLCFVHT	480		
OY	480 VPMDOLEFRNPQALHTANRPDE-CVGEGLACHOICARGHCGWPBPCTCVNCOSQFLRGO	538		
Dd	481 VPMDOLEFRNPQALHTANSRPREDECSSGIVONSLSIACGHCMGPPPTCVNCSHFLRGO	540		
OY	539 ECVEECRVLYGLPREYVNARHCLPCHPECOPONGSVTCFGEPAODCVACAHHKDPPECYA	598		
Dd	541 ECVEECRVWKGLPREYVSNKRCLPCHPECOPONSSETCFGESEADCAAHKHDSSSCYA	600		
OY	599 RCPSGVKPDLSTMPITPKFPDEBGACOPCINTCHSCVDDDKGCPAEQAASPLT-----	652		
Dd	601 RCPSGVKPDLSTMPITPKFYDEBGICOPCINTCHSCVDDDERGCPAEQAASPVTFIIATV	660		
OY	653 -----	652		
Dd	661 VGVLFLILYYVYGILIKRRROKIRKYTKRLLQETELVEPLTPSGAMENQOMRIKET	720		
OY	653 -----	652		
Dd	721 ELRKVKVLGSGAFGYVKKIWIIPDGENVKIIPAIVKLRENTSPKANKEILLDEAYVVAAGV	780		
OY	653 -----	652		

Dh	781	SPVYSRLIGLCTSTVYQVLVQMLPYGCLLDHVREHGRGLSGODLLNMCVQLAKGMSYLED	840
Qy	653	-----	652
Dh	841	VRLVHRDLAARNVLYKSPNHVKTDFGLARLLDIDETEYHADGCKVPITKMALESILRRR	900
Qy	653	-----	652
Dh	901	FTHOSDWMSYGYTWELMTFGAKPRYDGLIPAREIPDLLEKGERLPPQPICTIDVYIMVYKC	960
Qy	653	-----SONEDLGPASPLDSTFYRSLLEDMDGLV	682
Dh	961	WMIDSECRPRRELVEFSRMAKDFQREVVYVIONEDLGPSSPMDSFYRSLLEDMDGLV	1020
Qy	683	DAEELVYQOGGFEPCDDPAPGAGWVHHHRSSSTRSGGDLTLGLPESEEARSPPLAPS	742
Dh	1021	DAEELVYQOGGFEPCDDPAPGAGWVHHHRSSSTRSGGDLTLGLPESEEARSPPLAPS	1080
Qy	743	EGAGSDVDFDGLGMAAKGLQSLPTHDPSPLOKRYSEDETVPLPSETDGYVAPLTCSPQPE	802
Dh	1081	EGAGSDVDFDGLGMAAKGLQSLPTHDPSPLOKRYSEDETVPLPSETDGYVAPLTCSPQPE	1140
Qy	803	YVNGPDVAPRQPPSPPEGLPLAARPGATLERKTIISPKNGVYKRVYFAFGAVENPEYLT	862
Dh	1141	YVNGPDVAPRQPPSPPEGLPLAARPGATLERKTIISPKNGVYKRVYFAFGAVENPEYLT	1200
Qy	863	POGGAAPRPHPPAPSPAFNDLYYNDODPPEGAPSPFFKGTAEENEYGLIDVPPV	919
Dh	1201	PREGASPPHSPAPSPAFNDLYYNDODPPEGAPSPFFKGTAEENEYGLIDVPPV	1257
RESULT 3			
ID	ERB2_MESAU	STANDARD:	PRT: 1254 AA.
AC	Q60553:		
Dt	15-DEC-1998	(Rel. 37, Created)	
Dt	15-DEC-1998	(Rel. 37, Last sequence update)	
Dt	15-JUN-2002	(Rel. 41, Last annotation update)	
DE	Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)		
DE	(p185erbB2) (NEU proto-oncogene) (C-erbB-2).		
GN	ERBB2 OR NEU.		
OS	Mesocricetus auratus (Golden hamster).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;		
OC	Mesocricetus.		
OX	NCBI_TaxID=10036;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Nerve;		
RX	MEDLINE=94193007; PubMed=7908275;		
RA	Nakamura T., Ushijima T., Ishizaka Y., Nagao M., Arai M.,		
RA	Yamazaki Y., Ishikawa T.;		
RT	"Cloning and activation of the Syrian hamster neu proto-oncogene.;"		
RL	Gene 140:251-255(1994).		
CC	-I- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,		
CC	ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A		
CC	POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-		
CC	ALPHA AND AMPHIREGULIN (BY SIMILARITY).		
CC	-I- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein		
CC	tyrosine phosphate.		
CC	-I- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS		
CC	(POTENTIAL).		
CC	-I- SUBCELLULAR LOCATION: Type I membrane protein.		
CC	-I- PM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE		
CC	RESIDUES.		
CC	-I- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.		
CC	-----		
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; D16295; BAA03801.1; -.  
DR HSSP; P11362; IFGK.  
DR InterPro: IPR000494; EGFR\_L\_domain.  
DR InterPro: IPR000719; Euk\_pkinase.  
DR InterPro: IPR002174; Furin-like.  
DR InterPro: IPR001245; Tyr\_pkinase.  
DR InterPro: IPR004019; YLP\_motif.  
DR Pfam: PF00757; pkinase; 1.  
DR Pfam: PF01030; Furin-like; 1.  
DR Pfam: PF02757; YLP; 2.  
DR Pfam: PF02757; YLP; 2.  
DR ProDom: PD000001; Euk\_pkinase; 1.  
DR SMART; SM00261; FU; 3.  
DR SMART; SM00219; TYRK; 1.  
DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR; 1.  
DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;  
KW Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;  
KW Proto-oncogene; Disease mutation.  
FT SIGNAL 1 21  
FT CHAIN 22 1254  
FT DOMAIN 22 1254  
FT TRANSMEM 653 675  
FT DOMAIN 676 1254  
FT DOMAIN 158 368  
FT DOMAIN 472 644  
FT DOMAIN 720 734  
FT NP\_BIND 726 734  
FT BINDING 753 753  
FT ACT\_SITE 845 845  
FT DISULFID 195 204  
FT DISULFID 199 212  
FT DISULFID 236 244  
FT DISULFID 240 252  
FT DISULFID 255 264  
FT DISULFID 268 295  
FT DISULFID 299 311  
FT DISULFID 315 331  
FT DISULFID 334 338  
FT DISULFID 511 520  
FT DISULFID 515 528  
FT DISULFID 531 540  
FT DISULFID 544 560  
FT DISULFID 563 576  
FT DISULFID 567 584  
FT DISULFID 587 596  
FT DISULFID 600 623  
FT DISULFID 626 634  
FT DISULFID 630 642  
FT MOD\_RES 1139 1139  
FT MOD\_RES 1247 1247  
FT CARBOHYD 68 68  
FT CARBOHYD 125 125  
FT CARBOHYD 187 187  
FT CARBOHYD 259 259  
FT CARBOHYD 530 530  
FT CARBOHYD 571 571  
FT CARBOHYD 629 629  
FT VARIANT 658 658  
FT VARIANT 659 659  
SQ SEQUENCE 1254 AA; 138252 MW; 974C3791C21F2BE1 CRC64;

Query Match 81.2%; Score 4122.5; DB 1; Length 1254;  
Best local Similarity 61.8%; Pred. No. 9, 2e-231;  
Matches 776; Conservative 49; Mismatches 93; Indels 337; Gaps 2;

QY 1 MELALACRMGLLALLPPGAAS7OVTGTDMLRLPASPTHLMDLRLHLYOGGCVVQGNL 60  
DB 1 MELALACRMGLLALLPPGAAS7OVTGTDMLRLPASPTHLMDLRLHLYOGGCVVQGNL 60

QY 61 ELRYLPNASTLSPLODIOEVGYVYLIAHQVROVPLQRLRIVRGSTOLFEDNVALAVLDNG 120  
DB 61 ELRYLPNASTLSPLODIOEVGYVYLIAHQVROVPLQRLRIVRGSTOLFEDNVALAVLDNR 120  
QY 121 DPLNNTTPVVGASPGSGREVLQSLRSLTEILKGVLIQNNPOLCYODTITMIDIFKNNOLA 180  
DB 121 DPLDNTVTATGRPEGLREIQRLSLTEILKGVLIQNNPOLCYODTITMIDIFKNNOLA 180  
QY 181 LTLIDITNRSRACHPCSPMCKSGRCWGSSESDCSLRTVYAGGACGARKGPLPTDCHEQC 240  
DB 181 PVDIDITNRSRACHPCSPMCKSGRCWGSSESDCSLRTVYAGGACGARKGPLPTDCHEQC 240  
QY 241 AAGCTGKSHDCLACLAFLPNNSGICELCPALVYNTTFESSMPEBRYTFGASCTYACR 300  
DB 241 AAGCTGKSHDCLACLAFLPNNSGICELCPALVYNTTFESSMPEBRYTFGASCTYACR 300  
QY 301 YNYLSTVGSCTLYCPILHNOEVTAEADGTORCEKSKCAVYCGLGMEHLREYRAVTSAN 360  
DB 301 YNYLSTVGSCTLYCPILHNOEVTAEADGTORCEKSKCAVYCGLGMEHLREYRAVTSAN 360  
QY 361 IOEPAGCKITFGSLAFLPESFDGDPASNTAPLOPEOLQVEETLEETGYLYISAMPDLSL 420  
DB 361 IOEPAGCKITFGSLAFLPESFDGDPASNTAPLOPEOLQVEETLEETGYLYISAMPDLSL 420  
QY 421 DLSYFQNLQYIRGRILHNGAYSLTLOGLISWGLSKRLSGSLAIHNTILCYHTY 480  
DB 421 DLSYFQNLQYIRGRILHNGAYSLTLOGLISWGLSKRLSGSLAIHNTILCYHTY 480  
QY 481 PMDOLFNPQALHTANPREDECEVGEGLACHOLCAHGMGPGPTCVNCSOLFNRQEC 540  
DB 481 PMDOLFNPQALHTANPREDECEVGEGLACHOLCAHGMGPGPTCVNCSOLFNRQEC 540  
QY 541 VEECRVLQGLPREVYVNAHCLPCHPECPONGSVTCFEPADOCVACAHYKDDPFVCYAR 600  
DB 541 VEECRVLQGLPREVYVNAHCLPCHPECPONGSVTCFEPADOCVACAHYKDDPFVCYAR 600  
QY 601 PSQVKKPPLSTMPITMKPFDEGACQPCPINCTHSDYDLDKGCRAEONASLTS 653  
DB 601 PSQVKKPPLSTMPITMKPFDEGACQPCPINCTHSDYDLDKGCRAEONASLTS 653  
QY 654 ----- 653  
DB 654 ----- 653  
QY 661 ILFLVIGVYVGLIKRRRQKIRKYMRLLOTELVEPLTPSGAMPNQAOMLIKETEL 720  
DB 661 ILFLVIGVYVGLIKRRRQKIRKYMRLLOTELVEPLTPSGAMPNQAOMLIKETEL 720  
QY 721 RKVKVLGSAFGTYVKGKIWIIPGENVKIPVAIKVIRENTSPKANKETLDEAYVMAIGLSP 780  
DB 721 RKVKVLGSAFGTYVKGKIWIIPGENVKIPVAIKVIRENTSPKANKETLDEAYVMAIGLSP 780  
QY 781 YVSRLLGICLSTVOLVTOQLMPYGLLDVHREHGRGLSODLLMNCVOIAKMSYLEDVR 840  
DB 781 YVSRLLGICLSTVOLVTOQLMPYGLLDVHREHGRGLSODLLMNCVOIAKMSYLEDVR 840  
QY 841 LVHRDLAARVVLKSPNHVKITDEGLARLLDIDETEVHADGKGVPIKWLBSILRRRT 900  
DB 841 LVHRDLAARVVLKSPNHVKITDEGLARLLDIDETEVHADGKGVPIKWLBSILRRRT 900  
QY 901 HOSDVMSTGYVWELMTFGAKPYDIPAREIPDLLEKGERLPORICTIDVYIMVNCWM 960  
DB 901 HOSDVMSTGYVWELMTFGAKPYDIPAREIPDLLEKGERLPORICTIDVYIMVNCWM 960  
QY 961 IDSECRPRRELYSESRKARDPQREYVIONEDIGPSPLDSTFYSLDEDDMGDLVDA 1020  
DB 961 IDSECRPRRELYSESRKARDPQREYVIONEDIGPSPLDSTFYSLDEDDMGDLVDA 1020  
QY 1021 EBYLVPOQGFCDPAPAGAGVNHRRSSSTSGGDLTLGLPSESEAPRPLAPSEG 744  
DB 1021 EBYLVPOQGFCDPAPAGAGVNHRRSSSTSGGDLTLGLPSESEAPRPLAPSEG 744  
QY 1081 AGSDVFDGDLGKAAGLQSLPTHDSPLQRYSEDPVPLPSETDGVADLFTCSPOPEYV 804  
DB 1081 AGSDVFDGDLGKAAGLQSLPTHDSPLQRYSEDPVPLPSETDGVADLFTCSPOPEYV 804

OY 805 NOPDVRRPQPPSPREGPLPARPAGATLERPKTLSPGKNGVYKDVFAFGAVENPEYLTPQ 864  
 DB 1141 NOPEVRPQPLRPEGLRPPVRRGATLERPKTLSPGKNGVYKDVFFGAVENPEYLVR 1200  
 OY 865 GGAAPOPHPFPAFDFLXYWDDPPRGAPEPFSTFKTPTAENPEYGLVPPV 919  
 DB 1201 GGSASQPH-PPALCAFDNLYWDDPPSRGSPNTEFTPTAENPEYGLVPPV 1254

RESULT 4  
 EGFR\_HUMAN STANDARD: PRT: 1210 AA.  
 ID EGFR\_HUMAN  
 AC P00533; P06268; Q14225; Q9UMD7; Q9UMD8; Q9UMG5; Q9Z795; C00732;  
 AC 000688; Q9B2S2; Q9H2C9; Q9G2X1; Q9H3C9;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Receptor  
 protein-tyrosine kinase ErbB-1).  
 GN EGFR OR ERBB1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RX MEDLINE=84219729; PubMed=6328312;  
 RA Ullrich A., Coussens L., Hayflick J.S., Dull T.J., Gray A., Tam A.W.,  
 RA Lee J., Yarden Y., Libermann T.A., Schlessinger J., Downward J.,  
 RA Mayes E.L.V., Whittle N., Waterfield M.D., Seeburg P.H.;  
 RT "Human epidermal growth factor receptor cDNA sequence and aberrant  
 RT expression of the amplified gene in A431 epidermoid carcinoma cells.";  
 RT Nature 309:418-425(1984).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RX TISSUE=Placenta;  
 RX MEDLINE=95382857; PubMed=7654368;  
 RA Ilekis J.V., Stark B.C., Scoccia B.;  
 RT "Possible role of variant RNA transcripts in the regulation of  
 RT epidermal growth factor receptor expression in human placenta.";  
 RT Mol. Reprod. Dev. 41:149-156(1995).  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RX TISSUE=Placenta;  
 RX MEDLINE=97078686; PubMed=8918811;  
 RA Reltter J.L., Maible N.J.;  
 RT "A 1.8 kb alternative transcript from the human epidermal growth  
 RT factor receptor gene encodes a truncated form of the receptor.";  
 RT Nucleic Acids Res. 24:4050-4056(1996).  
 RN [4]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RX TISSUE=Placenta;  
 RX MEDLINE=97256547; PubMed=9103388;  
 RA Ilekis J.V., Garfili J., Niederberger C., Scoccia B.;  
 RT "Expression of a truncated epidermal growth factor receptor-like  
 RT protein (TEGFR) in ovarian cancer.";  
 RT Gynecol. Oncol. 65:36-41(1997).  
 RN [5]  
 RP SEQUENCE FROM N.A. (ISOFORMS 3 AND 4).  
 RX TISSUE=Placenta;  
 RX MEDLINE=21100872; PubMed=1161793;  
 RA Reltter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Daniels A.J.,  
 RA Schell Stinclair C., Pearls R.S., Green P.J., Yee D., Lampland A.L.,  
 RA Balasubramaniam S., Crossley T.D., Magnuson T.R., James C.D.,  
 RA Maible N.J.;  
 RT "Comparative genomic sequence analysis and isolation of human and  
 RT mouse alternative EGFR transcripts encoding truncated receptor  
 RT isoforms.";  
 RT Genomics 71:1-20(2001).  
 RN [6]  
 RP SEQUENCE OF 575-687 FROM N.A.  
 RA Reltter J.L., Threadgill D.W., Daniels A.J., Schell C.M.,  
 RA Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,

RA Maible N.J.;  
 RT "Human and mouse alternative EGFR transcripts encoding only the  
 RT extracellular domain of the receptor.";  
 RT Submitted (FEB-1999) to the EMBL/Genbank/DBJ databases.  
 RN [7]  
 RP SEQUENCE OF 713-924 FROM N.A.  
 RX MEDLINE=84196372; PubMed=6326261;  
 RA Lin C.R., Chen W.S., Krueger W., Stolarsky L.S., Weber W.,  
 RA Evans R.M., Verma I.M., Gill G.N., Rosenfeld M.G.;  
 RT "Expression cloning of human EGFR receptor complementary DNA: gene  
 RT amplification and three related messenger RNA products in A431  
 RT cells.";  
 RT Science 224:843-848(1984).  
 RN [8]  
 RP SEQUENCE OF 150-962 FROM N.A.  
 RX MEDLINE=84245835; PubMed=6330563;  
 RA Xu Y.H., Ishii S., Clark A.J.L., Sullivan M., Wilson R.K., Ma D.P.,  
 RA Roe B.A., Merlino G.T., Pastan I.;  
 RT "Human epidermal growth factor receptor cDNA is homologous to a  
 RT variety of RNAs overproduced in A431 carcinoma cells.";  
 RT Nature 309:806-810(1984).  
 RN [9]  
 RP SEQUENCE OF 1028-1210 FROM N.A.  
 RX MEDLINE=85046483; PubMed=6093780;  
 RA Simmen F.A., Gope M.L., Schultz T.Z., Wright D.A., Carpenter G.,  
 RA O'Malley B.W.;  
 RT "Isolation of an evolutionarily conserved epidermal growth factor  
 RT receptor cDNA from human A431 carcinoma cells.";  
 RT Biochem. Biophys. Res. Commun. 124:125-132(1984).  
 RN [10]  
 RP SEQUENCE OF 1-29 FROM N.A.  
 RX MEDLINE=88217333; PubMed=3329716;  
 RA Haley J.D., Whittle N., Bennett P., Kinchington D., Ullrich A.,  
 RA Waterfield M.D.;  
 RT "The human EGF receptor gene: structure of the 110 kb locus and  
 RT identification of sequences regulating its transcription.";  
 RT Oncogene Res. 1:375-396(1987).  
 RN [11]  
 RP SEQUENCE OF 1-29 FROM N.A.  
 RX MEDLINE=91107677; PubMed=1988448;  
 RA Haley J.D., Waterfield M.D.;  
 RT "Contributory effects of de novo transcription and premature  
 RT transcript termination in the regulation of human epidermal growth  
 RT factor receptor proto-oncogene RNA synthesis.";  
 RT J. Biol. Chem. 266:1746-1753(1991).  
 RN [12]  
 RP SEQUENCE OF 1-29 FROM N.A.  
 RX MEDLINE=85270438; PubMed=2991899;  
 RA Ishii S., Xu Y.H., Stratton R.H., Roe B.A., Merlino G.T., Pastan I.;  
 RT "Characterization and sequence of the promoter region of the human  
 RT epidermal growth factor receptor gene.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 82:4920-4924(1985).  
 RN [13]  
 RP SEQUENCE OF 540.  
 RA Kohda D.;  
 RT Submitted (SEP-1997) to the SWISS-PROT data bank.  
 RN [14]  
 RP RECEPTOR ACTIVITY.  
 RX MEDLINE=84191554; PubMed=6325948;  
 RA Mroczkowski B., Mosig G., Cohen S.;  
 RT "ATP-stimulated interaction between epidermal growth factor receptor  
 RT and supercoiled DNA.";  
 RT Nature 309:270-273(1984).  
 RN [15]  
 RP PHOSPHORYLATION.  
 RX MEDLINE=89278137; PubMed=2543678;  
 RA Margolis B.L., Lax I., Kris R., Dombalagian M., Honegger A.M.,  
 RA Howk R., Givol D., Ullrich A., Schlessinger J.;  
 RT "All autophosphorylation sites of epidermal growth factor (EGF)  
 RT receptor and HER2/neu are located in their carboxyl-terminal tails.  
 RT Identification of a novel site in EGF receptor.";  
 RT J. Biol. Chem. 264:10667-10671(1989).  
 RN [16]



RP CARBOHYDRATE-LINKAGE SITES ASN-128; ASN-175; ASN-413; ASN-444 AND  
 RP ASN-528  
 RX MEDLINE-96398132; PubMed-8962717;  
 RA Smith K.D., Davies M.J., Bailey D., Renouf D.V., Hounsell E.F.;  
 RT "Analysis of the glycosylation patterns of the extracellular domain of  
 the epidermal growth factor receptor expressed in Chinese hamster  
 RT ovary fibroblasts."  
 RL Growth Factors 13:121-132(1996).  
 RN [17]  
 RP CARBOHYDRATE-LINKAGE SITES ASN-56; ASN-352; ASN-361; ASN-568 AND  
 RP ASN-603.  
 RX MEDLINE-20198209; PubMed-10731668;  
 RA Sato C., Kim J.-H., Abe Y., Saito K., Yokoyama S., Kohda D.;  
 RT "Characterization of the N-oligosaccharides attached to the atypical  
 RT Asn-X-Cys sequence of recombinant human epidermal growth factor  
 RT receptor."  
 RL J. Biochem. 127:65-72(2000).  
 RN [18]  
 RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.  
 RX MEDLINE-98225196; PubMed-9556602;  
 RA Abe Y., Otake M., Inagaki F., Iax I., Schlessinger J., Kohda D.;  
 RT "Disulfide bond structure of human epidermal growth factor receptor";  
 RL J. Biol. Chem. 273:11150-11157(1998).  
 RN [19]  
 RP REVIEW.  
 RX MEDLINE-87297456; PubMed-3039909;  
 RA Carpenter G.;  
 RT "Receptors for epidermal growth factor and other polypeptide  
 RT mitogens."  
 RL Annu. Rev. Biochem. 56:881-914(1987).  
 CC -1- FUNCTION: Receptor for EGF, but also for other members of the EGF  
 CC family, as TGF-alpha, amphiregulin, betacellulin, heparin-binding  
 CC EGF-like growth factor, Gp30 and vaccinia virus growth factor. Is  
 CC involved in the control of cell growth and differentiation.  
 CC -1- FUNCTION: Isoform 2/truncated isoform may act as an antagonist.  
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine - ADP + protein  
 CC tyrosine phosphate.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Isoform 2 is  
 CC secreted.  
 CC -1- ALTERNATIVE PRODUCTS: 4 isoforms: 1/p170 (shown here), 2/p60/  
 CC truncated isoform/TEGFR, 3/p110 and 4; are produced by  
 CC alternative splicing.  
 CC -1- TISSUE SPECIFICITY: Expressed in placenta. Isoform 2 is also  
 CC expressed in ovarian cancers.  
 CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to  
 CC dimerization, internalization of the EGF-receptor complex,  
 CC induction of the tyrosine kinase activity, stimulation of cell DNA  
 CC synthesis, and cell proliferation.  
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: X00588; CAA25240.1; -;  
 DR EMBL: 095089; AAB53063.1; -;  
 DR EMBL: 048722; AAC50802.1; -;  
 DR EMBL: 048723; AAC50804.1; -;  
 DR EMBL: 048724; AAC50796.1; -;  
 DR EMBL: 048725; AAC50797.1; -;  
 DR EMBL: 048726; AAC50798.1; -;  
 CC  
 Query Match 32.4%; Score 1643; DB 1; Length 1210;  
 Best Local Similarity 29.6%; Pred. No. 1,2e-87;  
 Matches 374; Conservative 139; Mismatches 310; Indels 442; Gaps 22;

QY 11 LLLALLPFGCA--STQVCTGTDMLKRLPASPTLHMLRLHLYGCGGVVQGNLELYLPNN 68  
 DB 14 LLAALCPASRALEKRVKCGTGNKLTQLGTFEDHFLSLQRMFNCEVVLGNLEITYVQGN 73

QY 69 ASLSFLQDIQVQGYVLIHNOYRQVPLQRLRVKRGTLFEDNYALAYLDNGDPLNNTTP 128  
 DB 74 YDLSPFKTQIEVAGVYLILNIVVERLPLENLIIIRNMVENSALAYLNDY----- 126  
 QY 129 VNGASPGRLREIQLRSLTELKGGVLIQRNPOLCYODTLMMDFIKNNQLTLIDTR 188  
 DB 127 ---ANKTGLKELPMRLQELLGAVAFSNNPALCNVESTIQMRDIYSDLSNMSDFQNH 183  
 QY 189 SRAPCHSPCKGSRGRCWGSSESDQSLTRTVAGGCA-RCKGPLPTCCHEQACAGCTGP 247  
 DB 184 LSCQCKDPSFCNPGSCWGAEGEENQCKLTITCAQCSGRCRGPSPDCCHNCAGACTGP 243  
 QY 248 KHSDDLACLHFNHSGICEALCALYTNNDTESMNPGRRTFGACVYTAQYNTLSD 307  
 DB 244 RSDSCIVCKRFDEACVCKTCEPLMLYNPTQMDNPEGKYSFGATCVKCKPRNVVVD 303  
 QY 308 VGSCTLVCPHLHNOEYVAEDGTORCEKSRPCARVCYGLGMEHLREYRAVTSANIOEFGC 367  
 DB 304 HSCVACAGADSEYM-EEDGVKCKKCEGPCRKYCNIGTIGEFKDSLINATINIKFKNC 362  
 QY 368 KRIFGSLAPLPSFDDPASNNTAPLQPEQLQVETLEETIGLYLISWPDLSPLSVFQ 427  
 DB 363 TESISDHLIPLVAFRGDSFTHTPLDPQELDIKTVEITGFLIQAWPENRTDLHAFEN 422  
 QY 428 LOYIRGRILHNGAYSLTLOGLGISWGLSLBELSGSLIHNHTLSCVHTVPMQDLR 487  
 DB 423 LEIRKRTQHOQFSLAVVNLITSGLRSLKEISGDVITSGNKLCAVNTIMMKRLTG 482  
 QY 488 NPHQALLHPANPEDECVGEGLAGHOLCARGCMGPGPQVCNCSQFLRGQCEVERCYL 547  
 DB 483 TSGQKTIISNNGENCKATGQVCHALCSPEGMGPEPDCVSCRNVSGRECVKCNLL 542  
 QY 548 QGLPREYVNAHRCPLRPEPCQFQNSVTCGEPADQCVACAHKDPFPCVAPRSGVDP 607  
 DB 543 ECEPREFVENSECICQHPCLPLQAMNITCTGRGPNICQAHYDGPCHVKCPAGVME 602  
 QY 608 LSYMTPKRPDEGACQRPINCTHSCVDLDDKCGCA----- 644  
 DB 603 NNTL-VKRYADAGHVCCHLCPMCTYGTCTGPGLEGCTGPKIPSTATGMVALLLVVA 661  
 QY 645 -----DQRASPLTSQNE----- 656  
 DB 662 LGIGLFMRNRHIVKRTLRLLQERLVEPLTPSGAPRQALLRIKTEFEKKIVLGS 721  
 QY 657 ----- 656  
 DB 722 AFGTVYKGLMPEGEKVIPIVAIKELREATSPKANKELDEAYVNASVDNPHVCRLIGIC 781  
 QY 657 ----- 656  
 DB 782 LRTVQLITQLMPFGCLLDYVREHKNIGSYLLMWCVOIAKGMVLEDRRLVHRDLAAR 841  
 QY 657 -----DLG----- 659  
 DB 842 NVLYVTRQHVKTTPDGLALLGAEKEYNABEGKVPDKMALESILHRTYTHQSDVWYSG 901  
 QY 660 -----PA----- 661  
 DB 902 VIVWELMFGSKPYDGIPIASEIISILEKGERLPPRPICIDVYMIWVKCMIDADSRPKF 961  
 QY 662 -----SPDSTFYRSLDDDDMGDLVDAEVLVPOQ 692  
 DB 962 RELIIEFSKMAIDPORVLYIQDERMHLRSPDSINRGLMDEEDDDVVDVDEYLIPOQ 1021  
 QY 693 GFECDDPAPAGAGVNHHRSSSTFGGDLTLGLEPSEEAAPRSBLAPSEAGSDVFDG 752  
 DB 1022 GFF-----SSPSSSRPILSSLSATSN--NS 1045  
 QY 753 DVGGAAGLQSLPTHDSPLORYSEDPTVLPSET--DGYVAPLTCSPOPEYVNOQDVR 810  
 DB 1046 TVACIDRNQLQSCPIKEDSFQRYSSDPGALTEDSIDTFV-----FVPEYIND----- 1095

```

QY 811 POPSPEBGLPAPAPGATL-----ERPKTLSPGKNGVKDFAFGAVENPEYL-TPQ 864
DB 1096 -----SVP-KRPAPSVONPYVHNQPLNAPSROPHYOD--PHSTAVGNPEYLTVQ 1143
QY 865 GGAAPQHPAPAPSPADNLYMDO-----DP-----PERGAPPTFGTPTAAE 908
DB 1144 -----PTCVNSTPDSAPHAQKSHQISLDNPDYQODPFPEAKNGCIFKGS-TAE 1193
QY 909 NPEYL 913
DB 1194 NAEYL 1198

RESULT 5
EGFR_MOUSE
ID EGFR_MOUSE STANDARD; PRT; 1210 AA.
AC 001279;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112).
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C; TISSUE=Liver;
RX MEDLINE=93026370; PubMed=1408137;
RA Ayali A., Skorecki K., Yayon A., Givol D.;
RT "Promoter region of the murine fibroblast growth factor receptor 2
RT (bek/KGFR) gene."
RL Oncogene 7:1957-1962(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C, and CD-1; TISSUE=Liver, and Decidua;
RX MEDLINE=93126380; PubMed=7678348;
RA Paria B.C., Das S.K., Andrews G.K., Dey S.K.;
RT "Expression of the epidermal growth factor receptor gene is regulated
RT in mouse blastocysts during delayed implantation."
RL Proc. Natl. Acad. Sci. U.S.A. 90:55-59(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C; TISSUE=Liver;
RA Hibbs M.L.;
RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-B6/C3; TISSUE=Liver;
RX MEDLINE=94170986; PubMed=8125255;
RA Luettike N.C., Phillips H.K., Qiu T.H., Copeland N.G., Earp H.S.,
RA Jenkins N.A., Lee D.C.;
RT "The mouse waved-2 phenotype results from a point mutation in the EGF
RT receptor tyrosine kinase."
RL Genes Dev. 8:399-413(1994).
RN [5]
RP SEQUENCE OF 1-714 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=91232866; PubMed=2030916;
RA Ayali A., Lax I., Ullrich A., Schlessinger J., Givol D., Morse B.;
RT "Comparison of EGF receptor sequences as a guide to study the ligand
RT binding site."
RL Oncogene 6:673-676(1991).
RN [6]
RP SEQUENCE OF 969-1117 FROM N.A.
RC STRAIN=C3H;
RA Eslinger D.P., Serrero G.;
RL Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,
CC AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
CC VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine -> ADP + protein

```

```

CC tyrosine phosphate.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SUBCELLULARS: Binding of EGF to the receptor leads to
CC induction, internalization of the EGF-receptor complex,
CC dimerization of the tyrosine kinase activity, stimulation of cell DNA
CC synthesis, and cell proliferation.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X78987; CAAS5587.1; -
CC EMBL; U03425; AAA17899.1; -
CC EMBL; X59698; CAA42219.1; -
CC EMBL; L06864; AAA53029.1; -
CC EMBL; Z12608; CAA78249.1; -
CC HSP; P11362; IFCG.
CC MGD; MGI:95294; Egfr.
CC InterPro: IPR000494; EGFR_L_domain.
CC InterPro: IPR000719; Euk_pkinase.
CC InterPro: IPR002174; Furin-like.
CC InterPro: IPR001245; Tyr_pkinase.
CC Pfam; PF00069; Pkinase; 1.
CC Pfam; PF00757; Furin-like; 1.
CC Pfam; PF01030; Recep_L_domain; 2.
CC ProDom: PD000001; Euk_pkinase; 1.
CC SMART; SM00261; FU; 3.
CC SMART; SM00219; TYRKC; 1.
CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
CC PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
CC Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
CC Tyrosine-protein kinase; ATP-binding; Phosphorylation; Repeat.
CC SIGNAL 1 24
CC CHAIN 25 1210
CC DOMAIN 25 647
CC TRANSMEM 648 670
CC DOMAIN 671 1210
CC REPEAT 75 300
CC REPEAT 390 600
CC DOMAIN 1028 1071
CC DOMAIN 714 981
CC NP_BIND 720 728
CC BINDING 747 747
CC ACT_SITE 839 839
CC DISULFID 190 199
CC DISULFID 194 207
CC DISULFID 215 223
CC DISULFID 219 231
CC DISULFID 232 240
CC DISULFID 236 248
CC DISULFID 251 260
CC DISULFID 264 291
CC DISULFID 295 307
CC DISULFID 311 326
CC DISULFID 329 333
CC DISULFID 506 515
CC DISULFID 510 523
CC DISULFID 526 535
CC DISULFID 529 555
CC DISULFID 539 555
CC DISULFID 558 571
CC DISULFID 562 579
CC DISULFID 582 591
CC DISULFID 595 617
CC DISULFID 620 628
CC DISULFID 624 636
CC MOD_RES 680 680
CC MOD_RES 1092 1092
CC -----
CC PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
CC PHOSPHORYLATION (AUTO-) (BY SIMILARITY).

```

```
FT MOD_RES 1110 1110 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1172 1172 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1197 1197 PHOSPHORYLATION (AUTO-, MAJOR SITE)
FT CARBOHYD 128 128 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 175 175 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 352 352 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 413 413 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 444 444 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 528 528 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 568 568 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 603 603 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 19 19 C -> S (IN REF. 2).
FT CONFLICT 539 539 C -> W (IN REF. 5).
FT CONFLICT 991 991 L -> F (IN REF. 4).
FT CONFLICT 1116 1117 HP -> DR (IN REF. 6).
SQ SEQUENCE 1210 AA: 134853 MW: 690E20D46DFD2D2F5 CRC64;

Query Match 31.7% Score 1612; DB 1; Length 1210;
Best Local Similarity 29.6%; Pred. No. 7.7e-86;
Matches 378; Conservative 131; Mismatches 313; Indels 454; Gaps 25;

OY 11 LLLALLPAGAA--STOVCTGDMKLRIPASEPETHLDMRLHYOGCQVVOGNLELYLPTN 68
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 14 LLLTALCAGALAEKKKQVQGSNRLTQGTEDHFLSLQRMVNNCEVVLGNLEIYYORN 73
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 69 ASLSLADIOEVQGVVLIANHOVPLQRLRIYRGTOLEFDNVALAVLNDGPLNNTTP 128
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 74 YDLSELKTIQVAGVVLALTALVTVERIPLENQIIRGNALYENTVALIISN----- 124
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 129 VTGASPGGLRSLRSLRSLRSLRSLRSLRSLRSLRSLRSLRSLRSLRSLRSLRSLR 184
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 125 -YGNRKGRLRPLRNLOELILGAVRSNPILCMDTIYQNVNFMSSMDL--- 180
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 185 DTNSRACHPCSPMKSGSCWGESSEDCSLTRVYACAGCA-RCKGRLPTDCHEBOCAAG 243
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 181 -QSHSSGCRKDPSCSPNCSGCGGEGNQKLTIIICAQCCSHRCGRSPDCCCHQCAAG 239
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 244 CTGRKHSCLLCLHFNHSGICELHCPALVYNTDTFESMPNDEGRHYTFGASCVTAPRYN 303
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 240 CTGPRESDCLYCQKFODEATCKDCPRLMYNTYQMDVNBEGKYSFGATCVKCKCPRNY 299
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 304 LSTDVSGSTLVCPLHNDQVTAEDQORCEKSKPCARVCYGLGMHLEAVNATNSAOE 363
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 300 VVTDHSGVACRAGDPDYEV-BEDGIRKCKKCDGPCRKVCNGIGIEFKDTLSINATNKH 358
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 364 FAGCKIRGSLAFLPESFDPDPASTAPLQPEOLQVFTLEITGXYLIASMPDLPDLS 423
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 359 FKYCAIISGDHLILPVAKQDSFTTRPLDLRELEIKTVEITGFFLLIQAMPDMDLH 418
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 424 VFQNLQVIRGILHNGAVSLTQGLGISWGLRSLRELGSGLALIHNTHLCEVHTVWD 483
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 419 AFENLEIRGRKQHGQFSLAVVGLNTSLGLRSLSKEISDDVLIISGRNLCYANTIMWK 478
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 484 OLFRPHALHTANRPDEDEVGSLACHOICARGHCHGPRPTQCVNCSQLOEQECEVE 543
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 479 KLFGPNQKTKIMNRRARAKDKAVNHNVCNLSCESSCGMPRPDVCSONVSRRGECEK 538
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 544 CRVLQGLPREVYNARHCLPCHREPOGNSVTCFEPADQCVACAHYDPFCFVACRCSG 603
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 539 CNILEGEPREVENSECICQCHPECLPQAMNITTCGRGPDNCLQCAHYIDGHCYKTCAG 598
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 604 VKPDLSTYPIKFPDEBEGACOPCPINCTHSCVDLDDKGCRA----- 644
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 599 IMGENNTL-VMKYADANNVCHLCHANCTYGCAGPGLQCEVWPSGPKLPSTATIGVGLL 657
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 645 -----EQRASPLTSQNE----- 656
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 658 FIYVVALGIGLFMRRRIIVRRRLRLQLQERELVEPLTPSGEAPRNOAHLRIKTEFEKKI 717
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 657 ----- 656
```

```
DB 718 KVLGSGARGTYVKGKLMPEGEKVKLPVAKIKELREATSPKANKELDEAVYMASVDNPHVC 777
OY 657 ----- 656
DB 778 RLIGICLSTVOLITQMLPARGCLLDYVREKNDIGSQYLNMWQVQIAGKMYLEDRLVH 837
OY 657 -----DLG----- 659
DB 838 RDLAARNLVKTPQHVKTITDFGLAKLGAEEKYHAEGKYKPKMALESILHRYTHQS 897
OY 660 -----PA----- 661
DB 898 DWWSYGVTVWELMTEGSKPRYDGIPIRSDISLKEGERLPQRPICIDIVYIMVKOMTIDA 957
OY 662 -----SPDSTFYRSLDEDDMGDLVDAE 686
DB 958 DSRPKFRELILEFSKMARDPQRYLVTOGDERHMLPSPTDSNFYRALMDEEDMEDVDAD 1017
OY 687 YLVPQGFPCFPDPARGAGGVNHHRRSSSTRSGGDLTLGLEPSEEARSPRIAPSEGAG 746
DB 1018 YLIPQGF-----NSPT-----SRPPLSSLSAT 1043
OY 747 SDVEDGDLGMAKGLQSLPTHDPSPLQYSEDPTVPLPSET--DGYVAPRLTCSPOPEYV 804
DB 1044 SN-----NSTVACINRNGSCRVKEDAFLOQYSDPTGATVEDNIDAF-----PVPEYV 1093
OY 805 NQPDYRQPPSPRSPRLPAARACATL-----ERKTLSPGKNVYKQVFAFGCAVEPE 859
DB 1094 NQ-----SVF-KRPAGSVQNPVYHNQPLHAPGRDLHYQN--PHSNVGNPE 1137
OY 860 YL-FPGGGAAPRPHPRPFAFDNLVYWD-----DP-----PERGAPPSFEK 902
DB 1138 YLNTAQ-----PTLSSGFNSPALMIQKSHQSLMDNPDIQDFFPKETPNGLFK 1188
OY 903 GTPAENPEYLGLDVP 918
DB 1189 G-PTAENAEYLRAVP 1203

RESULT 6
ERB4_HUMAN
ID ERB4_HUMAN STANDARD; PRT; 1308 AA.
AC Q15303;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112)
DE (p180erbB4) (Tyrosine kinase-type cell surface receptor HER4).
GN ERB4 OR HER4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM JM-A).
RC TISSUE-Breast carcinoma;
RX MEDLINE=93189574; PubMed=8383326;
RA Plowman G.D., Culouscou J.-M., Whitney G.S., Green J.M., Carlton G.W.,
RA Foy L., Neubauer M.G., Shoyab M.;
RT "Ligand-specific activation of HER4/p180erbB4, a fourth member of the
RT epidermal growth factor receptor family.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:1746-1750(1993).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM JM-B).
RC TISSUE-Fetal brain;
RX MEDLINE=97476287; PubMed=9334263;
RA Elenius K., Cortas G., Paul S., Choi C.J., Rio C., Plowman G.D.,
RA Klagsdun M.;
RT "A novel juxtamembrane domain isoform of HER4/erbB4. Isoform-specific
RT tissue distribution and differential processing in response to
RT phorbol ester.";
RL J. Biol. Chem. 272:26761-26768(1997).
```



```

Db 536 CNLYDGEFFRENGSICVCECDPOCEKMEDELLTCHGPGFNCCKSFHDKPNCVCEKCD 595
QY 603 GVKPDLSTYPIKFPDEBACQPCPINCCHSCVDLDDKC-----PAEQ----- 647
Db 596 GAGANSF--IFRYADPDRECHPCNPCTGCGNPTSHDCIYYPWTGSHSTLQDHARTPLI 653
QY 648 -----ASPLTSQ----- 654
Db 654 AAGVIGLFLIVGLTFVAVYRRKSIKKKRALRFLTELVEPLPSPGTAPNOADRLI 713
QY 655 -----NEDLGP----- 660
Db 714 KETELKRVKVLGSGAFVYKGIWPEGETVKIPAKIKILNETTGPKANVEKDALIMA 773
QY 661 ----- 660
Db 774 SMDPHLVLLGCVLSPTQLVTLQMPHCCLLEYVHEKDNIGSOLLNMCVOIAKGMX 833
QY 661 ----- 660
Db 834 LEERRLVHDLAARNVLYKSPNHVKITDFGLARLEGEDEKEYNADGKMPKIMALECIH 893
QY 661 ----- 660
Db 894 YRKFTHSDVMSYGVTLWELMTEFGCKPYDGIPTREIPDLLEKGERLPQPPICIDVYVM 953
QY 661 -----ASPLDSTFYISLEDDM 678
Db 954 VKCMIDADSRRPKELAEFSRMARDPQRYLVIGDDRMKLPSPNDSKFFQNLDEEDL 1013
QY 679 GDLVDAEEYLPQOGFPCDPAPAGAGVHHHRSSSTFGSDLLTGLSESEAPRS- 737
Db 1014 EDMDAEELVP-QAFNIPP-----ITSTRARIDSNS-----EIGSPPRATYPMG 1061
QY 738 -----PLAP-SEGASDVFDGDLGMAKAGLSQSLPTHD 769
Db 1062 NQFYVRDGGFAAEGQSVYRAPRTSTIPAPVAAGATATIFDSCNGTLRKRVANHYE 1121
QY 770 PSPLDRTSDPTVPLPS-----ETDGYVAPLTCSPOPEYVNPQDVVRQPSPREGRIP 822
Db 1122 DSTRTYSDPTVFAPEPERSRGELDEGYTMPMRKPKOEYLNPAVE-----ENP 1172
QY 823 AARPGATLERPKLTSPGKNGVVKDVFARFAGAVENPEYTLPGGAAPQHPPPA----- 876
Db 1173 SKR-----KNGDLQ-----ALDNPETYNASNG-----PKAEDEYVN 1204
QY 877 -----ESPFDNLXYWDODPERGA--PSTFKGPT 906
Db 1205 EPLYLNTFANTLGKAEYLLKNLITLSPERAKKAFDNDVNNHSLPRSTLQHPDYLGEST 1264
QY 907 -----AENPEYL 913
Db 1265 KYFYKONGRIRPIVAENPEYL 1285

```

```

RA Marchionni M.A., Kelly R.A.;
RT "Neuregulins promote survival and growth of cardiac myocytes."
RT Persistence of ErbB2 and ErbB4 expression in neonatal and adult
RT ventricular myocytes."
RT J. Biol. Chem. 273:10261-10269(1998).
RN [2]
RP SEQUENCE OF 848-901 FROM N.A.
RC TISSUE-sciatic nerve;
RX MEDLINE-9122560; PubMed-2025425;
RA Lai C., Lemke G.;
RT "An extended family of protein-tyrosine kinase genes differentially
RT expressed in the vertebrate nervous system";
RL Neuron 6:691-704(1991).
RN [3]
RP SEQUENCE OF 1031-1198 FROM N.A.
RC STRAIN-Sprague-Dawley; TISSUE-Spinal cord;
RX MEDLINE-97184212; PubMed-9030624;
RA Carroll S.L., Miller M.L., Frohnert P.W., Kim S.S., Corbett J.A.;
RT "Expression of neuregulins and their putative receptors, ErbB2 and
RT ErbB3, is induced during Wallerian degeneration.";
RL J. Neurosci. 17:1642-1659(1997).
CC -1- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-
CC 2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
CC NTAK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
CC NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
CC RECEPTORS (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN THE DEVELOPING
CC NERVOUS SYSTEM. EXHIBITS DISTINCT AND HIGHLY REGIONALIZED PATTERNS
CC OF EXPRESSION IN THE ADULT BRAIN, WHERE IT IS MAINLY FOUND IN THE
CC RETICULAR NUCLEUS OF THE THALAMUS. VERY LOW LEVELS IN KIDNEY, AND
CC HEART.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb-sib.ch).
CC -----
DR EMBL: AF041838; AAC08899.1; -.
DR EMBL: U52531; AAC53051.1; -.
DR HSSP: P11362; 1RKG.
DR InterPro: IPR000494; EGFR_L_domain.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR001245; Tyr_pkinase.
DR InterPro: IPR004019; YLP_motif.
DR Pfam: PF00757; Furin-like; 1.
DR Pfam: PF00069; pkinase; 1.
DR Pfam: PF01030; Recep_L_domain; 2.
DR Pfam: PF02757; YLP; 2.
DR PRINTS: PR00109; TYRKINASE.
DR PRODOM: PD000001; Euk_pkinase; 1.
DR SMART: SM00261; FU; 4.
DR SMART: SM00219; TYRC; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT SIGNAL 1 25
FT CHAIN 26 1308 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-4.
FT DOMAIN 26 651 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 652 675 POTENTIAL.
FT DOMAIN 676 1308 CYTOPLASMIC (POTENTIAL).

```

```
FT DOMAIN 186 334 CYS-RICH.
FT DOMAIN 496 633 CYS-RICH.
FT DOMAIN 718 985 PROTEIN KINASE.
FT NP_BIND 724 732 ATP (BY SIMILARITY).
FT BINDING 751 751 ATP (BY SIMILARITY).
FT ACT_SITE 843 843 BY SIMILARITY.
FT DISULFID 189 197 BY SIMILARITY.
FT DISULFID 193 205 BY SIMILARITY.
FT DISULFID 213 221 BY SIMILARITY.
FT DISULFID 217 229 BY SIMILARITY.
FT DISULFID 230 238 BY SIMILARITY.
FT DISULFID 234 246 BY SIMILARITY.
FT DISULFID 249 258 BY SIMILARITY.
FT DISULFID 262 289 BY SIMILARITY.
FT DISULFID 293 304 BY SIMILARITY.
FT DISULFID 306 323 BY SIMILARITY.
FT DISULFID 326 330 BY SIMILARITY.
FT DISULFID 503 512 BY SIMILARITY.
FT DISULFID 507 520 BY SIMILARITY.
FT DISULFID 523 532 BY SIMILARITY.
FT DISULFID 536 552 BY SIMILARITY.
FT DISULFID 555 569 BY SIMILARITY.
FT DISULFID 559 577 BY SIMILARITY.
FT DISULFID 580 589 BY SIMILARITY.
FT DISULFID 593 614 BY SIMILARITY.
FT DISULFID 617 625 BY SIMILARITY.
FT DISULFID 621 633 BY SIMILARITY.
FT MOD_RES 1162 1162 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1188 1188 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1258 1258 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1284 1284 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 138 138 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 174 174 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 253 253 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 358 358 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 410 410 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 473 473 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 495 495 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 548 548 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 620 620 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 1062 1062 S -> N (IN REF. 3).
FT CONFLICT 1080 1082 PYT -> SYR (IN REF. 3).
SQ SEQUENCE 1308 AA; 146957 MW; D944B0996A0B41 CRC64;
```

Query Match 29.9%; Score 1519; DB 1; Length 1308;  
Best Local Similarity 28.1%; Pred. No. 2e-80;  
Matches 371; Conservative 143; Mismatches 364; Indels 442; Gaps 25;

```
QY 1 MELA-ALCRMGILL-ALLPPGASTOVCTGDMKRLRLPASPTHLDMLRHLVYOGCQVQ 57
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1 MKLATGLMWGSLVLAARTVQPSASQSCAGTENKLSLSDLEOYRALRKYENCEVYM 60

QY 58 GNELETLPLTNASLSTIADIOEVGYLLAHNOVROPRLRLRYRGTOLEFEDNALAVL 117
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 61 GNEITITSIEHRDLSFLRSIREYGVYVALNFRPLYLEMLRIIRGKTLEDEYALAIIF 120

QY 118 DNGDPLNNTTPVGTGASPGGRLRLSTETELKGGVLIQRNPOLCYOOTILMKDIEHNN 177
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 121 LNYKKDGNF-----GIDELGKLNITELNGGVYDOKKLCYATHTIMODIVRNPW 171

QY 178 QALTLTLDITNRSRACHPCSPWCKSGRCGSESSDCOSTLRTVCAGGC-ARCGELPTGCC 236
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 172 PSNMTLVSTIGSSGCGRHKSCIG-RCMGPTENHCQILTRIVCAEDCGRGRTGYVSDCC 230

QY 237 HEQCAAGCTGPKHSDCLACLFHNSGICEHLCPALVTYNTDFESMPNDEGRYTFGASCV 296
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 231 HRECAAGCGSGPKDPCACMFNFDSGACVYCPQTFEYVNPPTFQLEHNFNAKYTGACVY 290

QY 297 TACGYNTLSTDVGSCITLVCPAHNEVTAEDCTGCEKSKCAVYVGLGHEHLREVRAY 356
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 291 KCPCHNFV-VDSSSVCVACPSKMEV-EENGIKMKCKPCTDICPRACDGIIGSLMSAQTV 348
```

```
QY 357 TSANIOEFAGCKITFGSLAFLEPSFDGDPASNTAPLQPEQLVEETLEITGYLISAMP 416
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 349 DSSNIDKRIKNTKNGNIIFLVTGIHDPYNAIDAIDPEKLNIVRTFAEITGFLINQWPF 408

QY 417 DLEPDLVSFQMLQYIRGILHNGAVSLTLOGLGISWGLRSLRELGSGLAIHHNTHICE 476
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 409 PNMDFSVFSMLVIGFVLSGSLILTKOOGITSLQFOSIKETISAGNITYTNSNICY 468

QY 477 VHTVPMQDLFNPQALHTNPDEDECVSGELCHOLCARGHMGCPPTQVCSOFLR 536
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 469 YHTIMVTTLESTVQRIYIRDNRAENCTABGMVCNHLCSNDGCMGEPDCLCRFSR 528

QY 537 GOECVEECRYLQGLPREVYNAHCLPCHRECP-ONGSVTCFGEADQCVACAHYKDPPE 595
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 529 GKICIESCNLYDGEFEREENGISICEQDSQCEKMEKDLITGHPGPDCTGCSHPKQPN 588

QY 596 CVARCPGKVDLSYMPIKFPDEBGACPPINCTHSCVDLDKGC-----PA 644
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 589 CVKCPDVLQAGNSF--TFKYADODRECHPCHPNCTGCGNSPTSHDCTIYYPWTGHTLPQ 646

QY 645 EQP-----ASPLTSQ----- 654
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 647 HAPPLIAGYIGGLFIVLMALFPAYVYVRKKSIAKKALRRLFETELVEPLTSGTAPN 706

QY 655 -----NEDLGP----- 660
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 707 QALRIKETELKRVKVLGSGAFGVYKGIWVPEGEYKIVAIKILNETGPRANVEFM 766

QY 661 ----- 660
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 767 DEALIMASVDHPHLVRLGVLCSPTQLVTLQMPHGCLLEYVHEBKDNIGSOLLNMCVQ 826

QY 661 ----- 660
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 827 IAKGMVTEERRLVHRDLAANVLKPSNHNKIIDFGLARLLEDEKERYNADGSKMPLK 886

QY 661 ----- 660
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 887 MALCIIHYRFTHQSDVSYGVTIWELMTFGKRPYDGIPTREIPDLLEKGRPLQPPICT 946

QY 661 -----ASPLDSTFYRS 671
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 947 IDVYIVMKCWMIDADSRRPKRELAERSMARDPORYLYIQGDDRMKLPSPNDSEFON 1006

QY 672 LLEDDMDGLVDAEIEYLVPQOQFPCDP-----APGAGGVH 708
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1007 LLEDEDLDMMDAEIYLP-QAFNIPPTIYSRRIDSNRSEIGHSPRPATTPSGSOFV 1065

QY 709 HHRSSSTRSGGDLTGLPSEEBEAPRSLAPSEAGSDVFDGLGGAAGLQSLPTH 768
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1066 YODGGAFAQQG---MPMYTATTSIPEAPVA--QGATAEMFDDSCCGTLRKPPVPHVQ 1120

QY 769 DSPFLQKSEDPYVPLPS-----ETDGYAALPCTSPQRYVQMDVPRPSPREGPL 821
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1121 EDSSTQRYSAAPTAFAPERNRAELDEGYMTPMHDKQOYLYLPPVENPEVSRKKMDL 1180

QY 822 PAAPRAGATLPRK-----TSPGKNGVKYDVE-----AFGAVENPEYLTPOGGAAP 869
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1181 QA-----LDNPEYHASSSPPK---AEDYVNEPLYLNTFNALGNAEYMKKSLISYP 1230

QY 870 QPHPPAPSPAFDMLYYWDODPERGA--PSTRTKGTPT-----AENPEYL 913
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1231 E-----KAKKAFDNDPDMWNHSLPPRSTLQHPDYLQEQESTKTFYKONGRIRPIVAENPEYL 1285
```

RESULT 8  
EGFR\_CHICK  
ID EGFR\_CHICK STANDARD; PRT; 703 AA.  
AC P13387;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Epidermal growth factor precursor (Ec 2.7.1.112) (CER)

FT	DISULFID	630	638	BY SIMILARITY.
FT	DISULFID	634	646	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	134	134	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	190	190	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	200	200	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	359	359	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	368	368	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	420	420	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	573	573	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	578	578	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	613	613	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	633	633	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	648	648	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	NON_TER	703	703	N-LINKED (GLCNAC. . .) (POTENTIAL).
SEQ	SEQUENCE	703 AA;	77427 MW; AEF2DE1B735A690 CRC64;	
Query March 29.7%; Score 1509.5; DB 1; Length 703;				
Best Local Similarity 44.2%; Pred. No. 3.5e-80;				
Matches 286; Conservative 103; Mismatches 235; Indels 23; Gaps 9				
QY	8	RWGLLLALPPGAA-----STOVCTGTDMLRLPASPETHLDMLRHLYOGCOVYGNLE	61	
DB	13	RGAANVLVLLILGVALCAVAEEKKQOGNNKTLTGIGHEDHTFSIQRMNNEEVLNLE	72	
QY	62	LTLYETNASLSFLDIOEGVGYLLAHNOVRLQRLRIYRGTLFEDYNALAVLNDGD	121	
DB	73	ITVEHNNDTLFELKTIQEGVGYLLAHNVDPILPENLIQINGVLYDNSFLAVLSNYH	132	
QY	122	PLNNTPTYGASPGSLREQLRSLEIILKGVLIQRNPOLCYODTILMKDIFHKNNQLAL	181	
DB	133	-MNRQ-----GLRELPMKRISLEIINGGVKISINPMKLMVDVLMWDIIDTSRK-PL	182	
QY	182	TLID-TNRSRACHPCSPMCKSGRCGSESDQSILRTYACAGCA-RCKGRLPTDSCHEO	239	
DB	183	TVLDPAUSLSSCPKRCPCNCTEDHCAGAEQONQTLTYICAQCGSGRCGKGYSPDCHNQ	242	
QY	240	CAAGCTGKRHSDCIACLFHNSHSGICELCRPALVYNTDTFESMPNDEGRHYTGASCVTAC	299	
DB	243	CAAGCTGPRESDCLACRFRDDATCKDPCPLVLYNPTTYQMDVNPDEKYSFGATCVREK	302	
QY	300	PYNTISTVSGCTLVCPRLHNOEYVTDGTQREKSKCARVCYGLGMHLEVAAYISA	359	
DB	303	PHNVYVYDHDGSCVRSCTNDTYEV-BENVYRCKCKCDGLCSKVCNSGIGELGIIINSTNT	361	
QY	360	NIOEPAGCKITFGSLAFIPESPDGPAPANTAPLOEOLQOVFEETLEIGYLIYASMPDL	419	
DB	362	NIDSFKNCTKINGDYSILPVAFLGDAFTKLPRLDPKIDYFETVKEISEFLLIQAMPDRA	421	
QY	420	PDLVFOMLOYIRGRILHNGAYSLTLQIGISIMWGLRSRLRELSGLALIHNTHLCEVHT	479	
DB	422	TDLVAFEWLEIRGRTRKHOGYSLAVVNLKTIQSLGRSLKETSIDGDIAIMKRNKCYADT	481	
QY	480	VPMQDLFNPMDALHTANRREDECVGGSLACHQICAGHCHMGPRPTQCVNCSQFLRQOE	539	
DB	482	MNMSLEFATQSQKTKRIILNRKNKNDTARHVCDPCLCSOVGCGPAPFCISFRFSRKE	541	
QY	540	CVEEBCRVLOGLAPREVANRHCILCPHECOPONG---SYTCFEPADOCVCAHYKDPFPC	596	
DB	542	CVKQCNITLQGPREFRERBSKCLPGISELYQNSTAYNTTCSFPGRDHCKAKHAFIDGRHC	601	
QY	597	VARCPGCVKPDLSYMPIMKFPDEBEGACOPPCINCTHSCVDDIDDKGCP	643	
DB	602	VKAPCGAVLGENDTL-VMKYADANNAVCOIChNPCNCRGCKGPELEBGP	647	
RESULT 9				
ERB3_HUMAN	STANDARD;	PRT; 1342 AA.		
AC	P21860;			
DT	01-MAY-1991 (Rel. 18, Created)			
DT	01-MAY-1991 (Rel. 18, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)			

DE (c-erbB3) (tyrosine kinase-type cell surface receptor HER3).  
GN ERBB3 OR HER3.  
OC Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90083234; Pubmed=2687875.  
RA Kraus M.H., Issing W., Miki T., Popescu N.C., Aaronson S.A.;  
RT "Isolation and characterization of ERBB3, a third member of the  
RT ERBB/epidermal growth factor receptor family: evidence for  
RT overexpression in a subset of human mammary tumors.";  
RL Proc. Natl. Acad. Sci. U.S.A. 86:9193-9197(1989).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90311312; Pubmed=2164210;  
RA Ploman G.D., Whitley G.S., Neubauer M.G., Green J.M., McDonald V.L.,  
RA Todaro G.J., Shoyab M.;  
RT "Molecular cloning and expression of an additional epidermal growth  
RT factor receptor-related gene.";  
RL Proc. Natl. Acad. Sci. U.S.A. 87:4905-4909(1990).  
RN [3]  
RP SEQUENCE FROM N.A. (SHORT FORM).  
RX TISSUE=Placenta;  
RX MEDLINE=93282822; Pubmed=7685162;  
RA Katoh M., Yazaki Y., Sugimura T., Terada M.;  
RT "c-erbB3 gene encodes secreted as well as transmembrane receptor  
RT tyrosine kinase.";  
RL Biochem. Biophys. Res. Commun. 192:1189-1197(1993).  
CC - FUNCTION: BINDS AND IS ACTIVATED BY NEURIGULINS AND NTRK.  
CC - CATALYTIC ACTIVITY: ATP + a protein tyrosine - ADP + protein  
CC tyrosine phosphate.  
CC - SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS  
CC (POTENTIAL).  
CC - SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG FORM) AND  
CC SECRETED (SHORT FORM).  
CC - ALTERNATIVE PRODUCTS: TWO FORMS. A LONG TRANSMEMBRANE FORM AND A  
CC SHORT SECRETED FORM, EXIST DUE TO ALTERNATIVE SPLICING.  
CC - TISSUE SPECIFICITY: EPITHELIAL TISSUES AND BRAIN.  
CC - DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE  
CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.  
CC - PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES  
CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF  
CC PHOSPHATIDYLINOSITOL 3-KINASE (BY SIMILARITY).  
CC - DISEASE: OVEREXPRESSED IN A SUBSET OF HUMAN MAMMARY TUMORS.  
CC - SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch))  
CC -----  
DR EMBL: M29366; AAA35790.1; -  
DR EMBL: M34309; AAA35979.1; -  
DR EMBL: S61953; AAB26935.1; -  
DR PIR: A36223; A36223.  
DR HSSP: P11362; 1FGK.  
DR Genew: HGNC:3431; ERBB3.  
DR MIM: 190151; -  
DR InterPro: IPR000494; EGFR\_L\_domain.  
DR InterPro: IPR000719; Euk\_pkinase.  
DR InterPro: IPR002174; Furin-like.  
DR InterPro: IPR001245; Tyr\_pkinase.  
DR Pfam: PF00069; pkinase; 1.  
DR Pfam: PF00757; Furin-like; 1.  
DR Pfam: PF01030; Recep\_L\_domain; 2.  
DR ProDom: PD000001; Euk\_pkinase; 1.  
DR SMART: SM00261; FU; 3.  
DR SMART: SM00219; TyrKc; 1.

DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; FALSE\_NEG.  
DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; FALSE\_NEG.  
DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;  
KW Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;  
KW Alternative splicing.  
FT SIGNAL 1  
FT CHAIN 19  
FT DOMAIN 20 1342  
FT TRANSMEM 20 643  
FT DOMAIN 644 664  
FT DOMAIN 665 1342  
FT DOMAIN 709 966  
FT NP\_BIND 715 723  
FT BINDING 742 742  
FT ACT\_SITE 834 834  
FT DISULFID 186 194  
FT DISULFID 190 202  
FT DISULFID 210 218  
FT DISULFID 214 226  
FT DISULFID 227 235  
FT DISULFID 231 243  
FT DISULFID 246 255  
FT DISULFID 259 286  
FT DISULFID 290 301  
FT DISULFID 305 320  
FT DISULFID 323 327  
FT DISULFID 500 509  
FT DISULFID 504 517  
FT DISULFID 520 529  
FT DISULFID 533 549  
FT DISULFID 552 565  
FT DISULFID 556 573  
FT DISULFID 576 585  
FT DISULFID 589 610  
FT DISULFID 613 621  
FT DISULFID 617 629  
FT CARBOHYD 126 126  
FT CARBOHYD 250 250  
FT CARBOHYD 353 353  
FT CARBOHYD 408 408  
FT CARBOHYD 414 414  
FT CARBOHYD 437 437  
FT CARBOHYD 469 469  
FT CARBOHYD 522 522  
FT CARBOHYD 566 566  
FT CARBOHYD 616 616  
FT VARSPLIC 141 183  
FT VARSPLIC 184 1342  
FT CONFLICT 560 560  
FT CONFLICT 1064 1064  
SQ SEQUENCE 1342 AA; 148097 MW; 7201E7F66CA374BD CRC64;  
Query Match 28.3%; Score 1435; DB 1; Length 1342;  
Best Local Similarity 43.6%; Pred. No. 1; 4e-75;  
Matches 278; Conservative 92; Mismatches 244; Indels 24; Gaps 12;  
QY 10 GLLALLPPGAA--STQVCTGDMKRLPASPEHLDMLRHLYOGGVGVNLELTLYPT 67  
DB 11 GLIFSLARGSEVGNQAVCPGLNLGLSVGDAENQYQTLRYKLYREVEVVMGLIEVLTGH 70  
QY 68 NASLSFLDIOIQEVOGVVLAHNOVROVPLQRIYRGTLQFEDNVALAVLGDGDLNNNT 127  
DB 71 MADLSLQWIREVTGVVLAHNEFSLPLPNLVYRGVGVYGGKRAIFEW-----LNYT 125  
QY 128 PYTGASPGGLRLQLRLSLTEILKGVLLQRPOLCYQDTILMKDFHKNQALATLIDTN 187  
DB 126 ----NSSHALRQLRLQLTEILISGVYTEKKNKLCMHMDIDRDIVRDRD---ALIVKD 178  
QY 188 RSRACHPCSPMKSGRCWGESSEDCQSLTRIVYACGGC-ARCKGRLPTDCHQCAAGCTG 246  
DB 179 NGRSCPCHVEYKG-RCWGPGESEDCQTLTKTICAPQCNHGCPNPNOCHDECAAGCGSG 237







```

FT CARBOHYD 356 356 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 365 365 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 398 398 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 501 501 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 621 621 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1167 AA; 129934 MW; 4793E0749DC1D55A CRC64;

Query Match 26.2%; Score 1330; DB 1; Length 1167;
Best Local Similarity 41.7%; Pred. No. 1.4e-69;
Matches 269; Conservative 93; Mismatches 259; Indels 24; Gaps 11;

QY 4 AALCRNGILLALPRGAASF---OYCTGDMKRLRLPASEFTLMDLRHLXGCGVQVGN 59
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db AALQ--LLVLVLSRCSTDPDRKVCQGSNMOTM---LDNHYLAKKKYSSCQVLEN 62

QY 60 LELLYLPTNALSFLQDIOEVGVYLAHQVQVPLQRLRYRGTLQFEDNALAVLDN 119
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db LELTYQENDLSFLQSIGQVGVYLAHNEVSTPLVNRLLRGQNLVYEGNTFLVMSN 122

QY 120 GDLPLANTPYTGASPGGLRELQRLSLTEILKGVLIQRPOLCYODTILMKDIFHKNNOL 179
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 123 YOK-NFSSP--DYQVGLKQLQSLNLTLEISGCVKSHNPLLCNVETIMMDIVDKTSNP 179

QY 180 ALFLIDTNSRACHPCSPMKSGKSGESEDOSLRTFYCAGGC-ARCKGPLETDCHE 238
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 180 TMLPLHAERKQCKQCDHCGVNSCWAPRGHCQKFTKLCAQCNRRCKPRPIDCCNE 239

QY 239 QCAAGCTGPKHSDCLALCNHNSIGELHCPALVYNTDTFFESMPNDEGRYTGASCYTA 298
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 240 HCAGGCTGPRATDCLACRPDNDGTCKDPCRPKITYDIYSHQVVDNPNIKYTGACVKE 299

QY 299 CPNNTSTVYGSTLYVCPHNDVYTAEDGTQREKSKPCARVCYGLGMEHLEPRAVYS 358
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 300 CPNNTVYTE-GACVRSKCSAGMLEVD-ENGKRSCKPCDGCPRKVCDEIGLSLNTIAVNS 357

QY 359 ANIOEAGCKKFGSLAFIPESPDPASNTAFLOPEQLQVFTLEITGYSLSMPDS 418
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 358 TNRSSNCTKINGDIIILNNSFEGDPHKKIGTMDPEHMLNTLVKEITYGLIIMMPEN 417

QY 419 LPDLSTFQNLQVIRGRILHNGAYS-LTLOGISWLGSLNSRLSGSLALIHNTFLCFY 477
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 418 MTLSTFQNLQVIRGRITTSRGSEFYVQVYRHLQMLGSLSKVSAQNTILKTLQRLA 477

QY 478 HTVPMQDLFRNPQALLHNPANRPDECEVGEGLACRHLGARGHCGPPTQVNCQSFRLG 537
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 478 NTIMRRLFRSEDSISLEYDART-----ENQTCNNECSDEGCGMGPITWCVSLHVDRC 530

QY 538 QCEVEECRYLQGLPREYVANAHLCPHPEQPOQNGSVTCGPREADOCVACAHKDPFCY 597
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 531 GRGVASCNLQGEPRRAQVQDGRVQCHQELVQDLSLTGCPAPACSSAHFQDPQCI 590

QY 598 ARCPGSKPDLSTYPIKTFPEDEGACOPCPINTCHSCVDLDKGC 642
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 591 PRCPHILIDGDTL-IMKYADKKMGQCPCHQNTQCGSGRLSGC 634

RESULT 12
ID EGFR.DROME STANDARD: PRT: 1426 AA.
AC P04412: 061601; Q9W2G0; P81868;
DT 13-AUG-1987 (rel. 05, Created)
DT 15-DEC-1998 (rel. 37, Last sequence update)
DT 16-OCT-2001 (rel. 40, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Egfr)
GN (Gutken receptor) (Tropo protein) (Drosophila relative of ErbB).
OS EGFR OR TOP OR C-ERBB OR DER OR CG10079.
OC Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_Taxid=7227;

```

```

RN [1]
RP SEQUENCE FROM N.A. (ISOFORM TYPES I AND II).
RX MEDLINE-94350209; PubMed-8070664;
RA Clifford R., Schubach T.;
RT "Molecular analysis of the Drosophila EGF receptor homolog reveals
RT that several genetically defined classes of alleles cluster in
RT subdomains of the receptor protein.";
RL genetics 137:531-550(1994).
RN [2]
RP REVISIONS.
RA Clifford R., Schubach T.;
RL Submitted (MAR-1998) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE-85124611; PubMed-2982499;
RA Livneh E., Glazer L., Segal D., Schlessinger J., Shilo B.-Z.;
RT "The Drosophila EGF receptor gene homolog: conservation of both
RT hormone binding and kinase domains.";
RL Cell 40:599-607(1985).
RN [4]
RP SEQUENCE FROM N.A., CHARACTERIZATION, AND ALTERNATIVE SPLICING.
RC STRAIN-Oregon-R; TISSUE-embryo;
RX MEDLINE-87002474; PubMed-3093080;
RA Schejter E.D., Segal D., Glazer L., Shilo B.-Z.;
RT "Alternative 5' exons and tissue-specific expression of the
RT Drosophila EGF receptor homolog transcripts.";
RL Cell 46:1091-1101(1986).
RN [5]
RP SEQUENCE FROM N.A. (ISOFORM TYPE I), TISSUE SPECIFICITY, AND MUTATION
RP ANALYSIS.
RX MEDLINE-99102120; PubMed-9882502;
RA Iesochin A.M., Yu S.-Y., Katz J., Baker N.E.;
RT "Several levels of EGF receptor signaling during photoreceptor
RT specification in wild-type, Ellipse, and null mutant Drosophila.";
RL Dev. Biol. 205:129-144(1999).
RN [6]
RP SEQUENCE FROM N.A. (ISOFORM TYPE I).
RC STRAIN-Berkeley;
RX MEDLINE-20196006; PubMed-10731132;
RA Adams M.D., Celnikker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton R.C., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abiril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodex A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mervulov G., Milshina N.V., Modyarty C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleib J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reibert K., Remington K., Saunders R.D.C., Scheler F., Shen H.,
RA Shue B.C., Sideri-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.-A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

```

RA Zhong X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 RL [17]  
 RP SEQUENCE OF 959-1078 FROM N.A.  
 RC STRAIN-Daekwanryeong;  
 RX MEDLINE=85137938; PubMed=2983232;  
 RA Wadsworth S.C., Vincent W.S. II, Bilocheau-Wentworth D.;  
 RT "A *Drosophila* genomic sequence with homology to human epidermal  
 RT growth factor receptor.";  
 RL Nature 314:178-180(1985).  
 RN [8]  
 RP SEQUENCE OF 1133-1137, 1155-1167 AND 1209-1216 FROM N.A., AND MUTATION  
 RP ANALYSIS.  
 RX MEDLINE=92038942; PubMed=1936959;  
 RA Raz E., Schejter E.D., Shilo B.Z.;  
 RT "Intallelic complementation among DER/fib alleles: implications for  
 RT the mechanism of signal transduction by receptor-tyrosine kinases.";  
 RL Genetics 129:191-201(1991).  
 RN [9]  
 RP REVIEW.  
 RX MEDLINE=97248481; PubMed=9094709;  
 RA Perrimon N., Perkins L.A.;  
 RT "There must be 50 ways to rule the signal: the case of the *Drosophila*  
 RT EGF receptor.";  
 RL Cell 89:13-16(1997).  
 CC -1- FUNCTION: BINDS TO FOUR LIGANDS: SPITZ, GURKEN, VEIN AND ARGOS,  
 CC WHICH IS AN ANTAGONIST. TRANSDUCES THE SIGNAL THROUGH THE RAS-RAF-  
 CC MAPK PATHWAY. INVOLVED IN A MYRIAD OF DEVELOPMENTAL DECISIONS.  
 CC CRITICAL FOR THE PROLIFERATION OF IMAGINAL TISSUES, AND FOR THE  
 CC DETERMINATION OF BOTH THE ANTERO-POSTERIOR AND DORSO-VENTRAL  
 CC POLARITIES OF THE OOCYTE. IN THE EMBRYO, PLAYS A ROLE IN THE  
 CC ESTABLISHMENT OF VENTRAL CELL FATES, MAINTENANCE OF AMNIOSEROSA  
 CC AND VENTRAL NEUROECTODERMAL CELLS, GERM BAND RETRACTION, CELL FATE  
 CC SPECIFICATION IN THE CENTRAL NERVOUS SYSTEM AND PRODUCTION OF  
 CC CUTICLE.  
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine - ADP + protein  
 CC tyrosine phosphate.  
 CC -1- SUBCELLULAR LOCATION: ISOFORMS TYPE I AND TYPE II ARE TYPE I  
 CC MEMBRANE PROTEINS. ISOFORM TYPE III IS A TYPE III MEMBRANE  
 CC PROTEIN.  
 CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS. TYPE I (SHOWN HERE), TYPE II AND  
 CC TYPE III ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- TISSUE SPECIFICITY: UBQUITOUSLY EXPRESSED IN EMBRYOS. IN LARVAE,  
 CC UNIFORM EXPRESSION IS SEEN IN WING DISKS, GENITAL DISK, ANLAGEN OF  
 CC TESTIS AND OVARY, AND BRAIN CORTEX. IN EYE-ANTENNA DISK, HIGHEST  
 CC EXPRESSION IS ANTERIOR TO MORPHOGENETIC FURROW, LEVELS REMAIN HIGH  
 CC IN PHOTORECEPTOR PRECURSOR CELLS. THIS PATTERN IS REVERSED IN  
 CC POSTERIOR EYE DISK. IN ADULTS EXPRESSION IS HIGH IN BRAIN CORTEX  
 CC AND THORACIC AND ABDOMINAL GANGLIA.  
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: AF052754; AAC08536.1; -  
 DR EMBL: AF052753; AAC08536.1; JOINED.  
 DR EMBL: AF052754; AAC08535.1; -  
 DR EMBL: AF052752; AAC08535.1; JOINED.  
 DR EMBL: K03054; AAA51462.1; -  
 DR EMBL: K03417; AAA51460.1; -  
 DR EMBL: K03416; AAA50965.1; -  
 DR EMBL: K03418; AAA51461.1; -  
 DR EMBL: AF109077; AAD26134.1; -  
 DR EMBL: AF109078; AAD26132.1; -  
 DR EMBL: AF109082; AAD26132.1; JOINED.  
 DR EMBL: AF109078; AAD26133.1; -  
 DR EMBL: AF109078; AAD26133.1; -

DR EMBL: AF109084; AAD26133.1; JOINED.  
 DR EMBL: AF109079; AAD26130.1; -  
 DR EMBL: AF109081; AAD26130.1; JOINED.  
 DR EMBL: AF109079; AAD26131.1; -  
 DR EMBL: AF109083; AAD26131.1; JOINED.  
 DR EMBL: AF109080; AAD26135.1; -  
 DR EMBL: AE003454; AAF46732.1; -  
 DR EMBL: X02293; CAA26157.1; -  
 DR EMBL: X78920; CAA55523.1; -  
 DR EMBL: X78918; CAA55521.1; -  
 DR EMBL: X78919; CAA55522.1; -  
 DR PIR: A00640; GOFEE.  
 DR HSP; P11362; IFEK.  
 DR FlyBase: FBgn0003731; Egr.  
 DR InterPro: IPR000494; EGFR\_L.domain.  
 DR InterPro: IPR000719; Euk\_kinase.  
 DR InterPro: IPR002174; Furin-kinase.  
 DR InterPro: IPR001245; Tyr\_kinase.  
 DR Pfam: PF00069; Kinase; 1.  
 DR Pfam: PF00757; Furin-like; 1.  
 DR Pfam: PF01030; Recep\_L.domain; 2.  
 DR PRINTS: PR00109; TYRKINASE.  
 DR ProDom: PD000001; Euk\_kinase; 1.  
 DR SMART: SM00261; FU; 7.  
 DR SMART: SM00219; Tyrc; 1.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR; 1.  
 DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR Transmembrane: Glycoprotein; Receptor; Phosphorylation; Transferase;  
 KW Tyrosine-protein kinase; ATP-binding; Signal; Alternative splicing;  
 KW developmental protein.  
 FT SIGNAL 1 30  
 FT CHAIN 31 1426  
 FT DOMAIN 31 868  
 FT TRANSMEM 869 889  
 FT DOMAIN 890 1426  
 FT DOMAIN 938 1196  
 FT NP\_BIND 944 952  
 FT BINDING 971 971  
 FT ACT\_SITE 1063 1063  
 FT MOD\_RES 902 902  
 FT PHOSPHORYLATION (BY PKC) (BY SIMILARITY).  
 Query Match 20.0%; Score 1013.5; DB 1; Length 1426;  
 Best Local Similarity 33.0%; Pred. No. 3.3e-51;  
 Matches 231; Conservative 93; Mismatches 280; Indels 97; Gaps 20;  
 OY 24 OVCSTDMKRLRPASPERHLMRLHYOCVOGNELETLPT-NASTSLFDIOEVOG 82  
 DB 100 KICITCKRLSLVSPSKKEHHYRLNDRYNTCTYVDGNLKLTPNPENLDSLFDINIREVTG 159  
 OY 83 YVLAHNOVROVPLQRLRIVRGTOLE-----EDNYALAVLDNGDPLNNTPTVTGASPGGL 137  
 DB 160 YILSHVDVKVKKPKQLIIRGRLFLSLVEEKALFV-----TYSKM 203  
 OY 138 RELQIRSLTEILKGVLIQIRNPOLCYDTILMKDIFKKNOLALTLDTNSRACHPCSP 197  
 DB 204 YLLEIPDLRDVLDNGVGHNNYNLCMHMTIOWSELVNGTDAYNYNFTAPERCPCPKE 263  
 OY 198 MCKGSRGCGESSEDDQSLRTVCAGGCA--CCKGPLPLPDCHEOCAGCCTPKKPSDCLAC 255  
 DB 264 SCTHGS-CWGEGRKCKQKSKLTCTSPQACAGRCYGGKPRECHLPCAGGCTGPTPKDCLAC 322  
 OY 256 LHFNSGICELHCPALVYNTDTEFSMPNPEGRTYFGASCYTACPYNLTSDVGSCTIYC 315  
 DB 323 KNFPENAVSKKECPMRKYNFTTYLLENPPGKAYAGTCKECP-GHLLDNGACVASC 381  
 OY 316 PLHNOEVTAEQTCRCEKSPCARVCYGLGMEHLREYRAVTSANIOEFACCKKIIFGSLA 375  
 DB 382 PODKMDKCGE-----CVPDNGPCPCTCPGYVLH-----AGNIDSFRCNCTYIDGNIR 428  
 OY 376 FLPESEFDG--DPASNTA-----PLQPBQLOVFEILEITTYIVTISAMPDSDPLSVYQN 427  
 DB 429 ILDTFSGFQDVYVANTYMGPRYIPLDPERREVFTVKEITTYLNIETGHPQFRMLSYFRN 488



```

FT DISULFID 670 693 POTENTIAL.
FT DISULFID 696 703 POTENTIAL.
FT DISULFID 700 715 POTENTIAL.
FT DISULFID 717 731 POTENTIAL.
FT DISULFID 735 750 POTENTIAL.
FT DISULFID 753 763 POTENTIAL.
FT DISULFID 757 771 POTENTIAL.
FT DISULFID 774 787 POTENTIAL.
FT DISULFID 791 805 POTENTIAL.
FT CARBOHYD 91 91 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 169 169 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 255 255 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 376 376 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 561 561 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 655 655 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 746 746 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 776 776 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 368 368 C -> Y (IN SY10).
FT VARIANT 469 469 C -> R (IN MN216).
FT VARIANT 700 700 C -> W (IN MN23).
FT VARIANT 753 753 C -> Y (IN SY11).
FT VARIANT 1065 1065 T -> I (IN SY16).
FT VARIANT 1074 1074 G -> E (IN SY7).
SQ SEQUENCE 1323 AA; 150510 MW; 6B0307E53BEF99 CRC64;

```

Query Match 14.1%; Score 718.5; DB 1; Length 1323;  
 Best Local Similarity 28.0%; Pred. No. 3.2e-34;  
 Matches 191; Conservative 107; Mismatches 256; Indels 129; Gaps 28;

```

OY . 25 VCTGDMKRLRASPETHLDMRLHLYGCGVQVQGNLETLPTN----- 68
Db . 39 LCGTGNGISRYGTNT-LEDELTMYRCGRRYGNYLLETMEANIKKWRSTSTVDPK 97
OY 69 -----ASLSFLQDIOEVGYVLIANOVROVPLRLIRVGTQLEFNVALAVLNDGDP 122
Db 98 NEDSPLKSLNFDNLEELRGLIIRANIKRSPRLRYGVGEVFNH-ALYIKHNK- 155
OY 123 LNNTPVTGASBGGLRELQSLSLTEILKGVLIQNPOLCY-QDTILMKDIFHKNNQAL 181
Db 156 -----VHEVYMRLEIRYRNGSVTIQNPCKYIGDKIDMKELLYDSD-VQ 199
OY 182 TLIDNRSRACH-----PCSPMKGSRGSESEDCOSLTRVCAAGGACRC---KGPL 231
Db 200 KVEITNSHQHCYONGKSMACHESC-NDKCMWSGNDQORYRSVCPSKSCSYFSTNSTS 258
OY 232 PTDCHEOCAGCTGPKHSDCLACLAHFNHSGICELHCPALVYNNDFESMPNPGRYTF 291
Db 259 SYECODSACLGCGTGHGPKNCKIACKYELDGCITCPSRKIFNFKTGLVYNPDPGRYON 318
OY 292 GASCTACPYNYL-STDVGSCTLYC-PLHNOEVTAEDESTORCEK-SKPCARVCYGLGME 348
Db 319 GNHCYKCEPPELLIENDV--CVRHCSGDGHYDATKD--VRECEKRSSSCPKICVVDG-- 372
OY 349 HLREVRATSNANIOEFACKKIFGSLAFLPESFODCPASNTAPLOPELOVFEETIELTG 408
Db 373 HL-----TNETLKNLECEQIDGHL-IIEHAF-----TYEOLKYLETVKIYSE 414
OY 409 YLYISAMPDLSLVSFONLOVIRGRILHNGAVSLTL-OGGISWLGKRSIRELSGSLAL 467
Db 415 YITIV--QONFYDLKFLKNLOIIEGRKLHNYRMALAIYQCDDELESLNKLITGAVL 472
OY 468 IHNHTHLCFTVPMDOLE-----RNPQOALLHTANRPDECVGEGELACHOLCARGCHGP 523
Db 473 IMKNHRLCYLSKIDMSIITSKGNKNSPLAIAENRSDKLETEBGRVCDKNKNGCGMK 532
OY 524 GPOTCVNCSQRLRGQCYEECRVLOGLPREYVNAHRCPLCHPECOQPONGSVTCFEPREADQ 583
Db 533 EPECELECKTWSVGTCEVKCDT-KGFLRNQTSMK-CERCSPECE-----TCNGLGELD 584
OY 584 CVACAH-----YKDPFCVARGCPSSGKPDLSYMPIMKPFDEBACQPCPINCINCHSCVD 636
Db 585 CLTGRHKLTIYNSDFGNRMCEVHDC-----PVSHFPTQKNVCEKCHPTCY----- 628

```

```

OY 637 LDDKGPAPQASPLTSONEDLG 659
Db 629 --DNCC-----TGPDNLNG 640

RESULT 14
ID ERB2_MOUSE STANDARD; PRT; 245 AA.
AC P70424; 061525;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-2 (EC 2.7.1.112) (p185erbB2)
DE (NEU proto-oncogene) (C-erbB-2) (Fragments).
GN ERBB2 OR NEU.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
(1)
RN SEQUENCE OF 1-149 FROM N.A.
RP STRAIN=CD-1; TISSUE=uterus;
RC MEDLINE=97200814; PubMed=9048643;
RX Lin J., Dey S.K., Das S.K.;
RA "Differential expression of the erbB2 gene in the perimplantation
RT mouse uterus: potential mediator of signaling by epidermal growth
RT factor-like growth factors."
RL Endocrinology 138:1328-1337(1997).
RN [2]
RP SEQUENCE OF 150-245 FROM N.A.
RX MEDLINE=96069911; PubMed=7589796;
RA Moscoso L.M., Chu G.C., Gautam M., Noakes P.G., Merlie J.P.,
RA Sanes J.R.;
RT "Synapse-associated expression of an acetylcholine receptor-inducing
RT protein, ARIA/heregulin, and its putative receptors, ErbB2 and ErbB3,
RT in developing mammalian muscle."
RL Dev. Biol. 172:158-169(1995).
CC -1- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC ALPHA AND AMPHIREGULIN.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine - ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN UTERINE EPITHELIAL
CC CELLS. IN THE MUSCLE, EXPRESSION LOCALIZES TO THE SYNAPTIC SITES
CC OF MUSCLE FIBERS.
CC -1- DEVELOPMENTAL STAGE: ON DAYS 1-4 OF PREGNANCY, ERBB2 IS DETECTED
CC PRIMARILY IN EPITHELIAL CELLS, THE DAY 1 UTERUS SHOWING THE
CC HIGHEST ACCUMULATION. ON DAY 5, THE EPITHELIUM AND THE
CC DECIDUALIZING SPROMAL CELLS AROUND THE IMPLANTING BLASTOCYST
CC EXHIBIT ACCUMULATION OF THIS RECEPTOR. ON DAYS 6-8, THE EXPRESSION
CC PERSISTS IN THE EPITHELIUM AT BOTH THE IMPLANTATION AND
CC SECONDARY DECIDUAL ZONE. ON DAYS 7 AND 8, ACCUMULATION IS ALSO
CC PROMINENT IN THE TROPHOBLASTIC GIANT CELLS.
CC -1- PM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES (BY STIMULATORY).
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U71126; AAB17380.1; -;
DR EMBL; L47239; AAA93532.1; -;
DR HSSP; P11362; IFGK.

```







GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 13, 2003, 14:44:10 ; Search time 9.60392 Seconds  
(without alignments)  
3074.904 Million cell updates/sec

Title: US-09-854-356-7  
3954

Sequence: 1 MELALCRWGLLALLRPGA.....GFCPPDPAPGAGVHHRRH 712

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3776	95.5	1255	1	ERB2_HUMAN
2	3205	81.1	1257	1	ERB2_RAT
3	3185	80.6	1254	1	ERB2_MESAU
4	1533.5	38.8	1210	1	EGFR_MOUSE
5	1532	38.7	1210	1	EGFR_HUMAN
6	1509.5	38.2	703	1	EGFR_CHICK
7	1451	36.7	1308	1	ERB4_HUMAN
8	1440.5	36.4	1308	1	ERB4_RAT
9	1435	36.3	1342	1	ERB3_HUMAN
10	1374.5	34.8	1339	1	ERB3_RAT
11	1330	33.6	1167	1	XMRK_XIPMA
12	1013.5	25.6	1426	1	EGFR_DROME
13	718.5	18.2	1323	1	LY23_CAEL
14	384	9.7	1363	1	ILPR_BRALA
15	383	9.7	1300	1	IRR_MOUSE
16	371	9.4	1300	1	IRR_CAVPO
17	368	9.3	1382	1	INSR_HUMAN
18	364	9.2	581	1	IRR_RAT
19	362.5	9.2	1383	1	INSR_RAT
20	361	9.1	1372	1	INSR_MOUSE
21	358	9.1	1297	1	IRR_HUMAN
22	353	8.9	1477	1	HTR7_HYDDE
23	326	8.2	1390	1	INSR_AEDAE
24	324	8.2	1607	1	MYR_LYMST
25	316	8.0	1370	1	IGIR_RAT
26	314	7.9	1373	1	IGIR_MOUSE
27	311.5	7.9	1367	1	IGIR_HUMAN
28	307	7.8	2146	1	INSR_DROME
29	271.5	6.9	1696	1	PCK5_MOUSE
30	244.5	6.2	1877	1	PCK5_MOUSE
31	208	5.3	1680	1	FUR2_DROME
32	195.5	4.9	937	1	PAC4_RAT
33	184.5	4.7	913	1	PCK5_HUMAN

34	181	4.6	245	1	ERB2_MOUSE	P70424 mus musculus
35	180.5	4.6	1877	1	PCK5_RAT	P41413 rattus norv
36	176	4.5	634	1	ERBB_ALV	P00534 avian leuco
37	176	4.5	969	1	PAC4_HUMAN	P29122 homo sapien
38	174	4.4	604	1	ERBB_AVIER	P00535 avian eryth
39	171.5	4.3	2003	1	NTCA_HUMAN	O99466 homo sapien
40	168.5	4.3	3110	1	LMG3_MOUSE	P24043 homo sapien
41	164.5	4.2	3695	1	LMAS_HUMAN	O15230 homo sapien
42	164	4.1	540	1	ERBB_AVIEU	P11273 avian eryth
43	162.5	4.1	1581	1	LMG3_MOUSE	O99066 mus musculu
44	162	4.1	1173	1	TSP1_XENLA	P35448 xenopus lae
45	161	4.1	3084	1	LMAL_MOUSE	P19137 mus musculu

## ALIGNMENTS

RESULT 1  
ID ERB2\_HUMAN STANDARD: PRT: 1255 AA.  
AC P04626:  
DR 13-AUG-1987 (Rel. 05, Created)  
DT 13-AUG-1987 (Rel. 05, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)  
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Tyrosine kinase-type cell  
DE surface receptor HER2) (MLN 19).  
GN ERBB2 OR HER2 OR NGL OR NEU.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Carnivora; Homiidae; Homo.  
OX NCBI\_Taxid:9606;  
[1]  
RX MEDLINE=86118663; PubMed=3003577;  
RA Yamamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N.,  
RA Saito T., Toyoshima K.,  
RT "Similarity of protein encoded by the human c-erb-B-2 gene to  
RT epidermal growth factor receptor.";  
RL Nature 319:230-234(1986).  
[2]  
RX MEDLINE=86070181; PubMed=2999974;  
RA Cousens L., Yang-Feng T.L., Liao Y.C., Chen E., Gray A.,  
RA McGrath J., Seeburg P.H., Libermann T.A., Schlessinger J.,  
RA Francke U., Levinson A., Ullrich A.,  
RT "Tyrosine kinase receptor with extensive homology to EGF receptor  
RT shares chromosomal location with neu oncogene.";  
RL Science 230:1132-1139(1985).  
[3]  
RX MEDLINE=737-1031 FROM N.A.  
RA MEDLINE=86016729; PubMed=2995967;  
RA Semba K., Kamata N., Toyoshima K., Yamamoto T.,  
RT "A v-erbB-related protooncogene, c-erbB-2, is distinct from the  
RT c-erbB-1/epidermal growth factor-receptor gene and is amplified in a  
RT human salivary gland adenocarcinoma.";  
RL Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501(1985).  
[4]  
RX MEDLINE=93194196; PubMed=8095488;  
RA Ehsani A., Low J., Wallace R.B., Wu A.M.,  
RT "Characterization of a new allele of the human ERBB2 gene by allele-  
RT specific competition hybridization.";  
RL Genomics 15:426-429(1993).  
-1- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,  
ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A  
POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-  
ALPHA AND AMPHIREGULIN.  
-1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
tyrosine phosphate.  
-1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS  
(POTENTIAL).  
-1- SUBCELLULAR LOCATION: Type I membrane protein.

CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE  
 CC RESIDUES (BY SIMILARITY).  
 CC -1- POLYMORPHISM: THERE ARE FOURS ALLELES DUE TO THE VARIATIONS IN  
 CC POSITIONS 654 AND 655. ALLELE B1 (654-ILE-ILE-655) HAS A FREQUENCY  
 CC OF 0.782; ALLELE B2 (654-ILE-VAL-655) HAS A FREQUENCY OF 0.206;  
 CC ALLELE B3 (654-VAL-VAL-655) HAS A FREQUENCY OF 0.012.  
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M11767; AAA35808.1; -.  
 DR EMBL: M11761; AAA35808.1; JOINED.  
 DR EMBL: M11762; AAA35808.1; JOINED.  
 DR EMBL: M11763; AAA35808.1; JOINED.  
 DR EMBL: M11764; AAA35808.1; JOINED.  
 DR EMBL: M11765; AAA35808.1; JOINED.  
 DR EMBL: M11766; AAA35808.1; JOINED.  
 DR EMBL: M11730; AAA75493.1; -.  
 DR EMBL: M12036; AAA35978.1; -.  
 DR EMBL: X03363; CAA27060.1; -.  
 DR PIR: A25491; A25491.  
 DR PIR: A24571; A24571.  
 DR HSSP: P11362; 1FGK.  
 DR Genew; HGNC:3430; ERBB2.  
 DR MIM: 164870; -.  
 DR InterPro: IPR000494; EGFR\_L\_domain.  
 DR InterPro: IPR000719; Euk\_kinase.  
 DR InterPro: IPR002174; Furin-like.  
 DR InterPro: IPR001245; Tyr\_kinase.  
 DR InterPro: IPR004019; YLP\_motif.  
 DR Pfam: PF00069; kinase; 1.  
 DR Pfam: PF00757; Furin-like; 1.  
 DR Pfam: PF01030; Recep\_L\_domain; 2.  
 DR Pfam: PF02757; YLP; 2.  
 DR Prodom: PD0000001; Euk\_kinase; 1.  
 DR SMART: SM00261; F0; 3.  
 DR SMART: SM00219; Tyrc; 1.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_APP; 1.  
 DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR; 1.  
 DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
 KM Transmembrane, Glycoprotein, Multigene family, Receptor, Signal;  
 KM Transferrin, Tyrosine-protein kinase; ATP-binding; Phosphorylation;  
 KM Polymorphism.  
 FT CHAIN 1 21 POTENTIAL.  
 FT CHAIN 22 1255 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.  
 FT DOMAIN 22 652 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 653 675 POTENTIAL.  
 FT DOMAIN 676 1255 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 720 987 PROTEIN KINASE.  
 FT NP\_BIND 726 734 ATP (BY SIMILARITY).  
 FT BINDING 753 753 ATP (BY SIMILARITY).  
 FT ACT\_SITE 845 845 BY SIMILARITY.  
 FT DISULFID 195 204 BY SIMILARITY.  
 FT DISULFID 199 212 BY SIMILARITY.  
 FT DISULFID 220 227 BY SIMILARITY.  
 FT DISULFID 224 235 BY SIMILARITY.  
 FT DISULFID 236 244 BY SIMILARITY.  
 FT DISULFID 240 252 BY SIMILARITY.  
 FT DISULFID 255 264 BY SIMILARITY.  
 FT DISULFID 268 295 BY SIMILARITY.  
 FT DISULFID 299 311 BY SIMILARITY.  
 FT DISULFID 315 331 BY SIMILARITY.  
 FT DISULFID 334 338 BY SIMILARITY.  
 FT DISULFID 511 520 BY SIMILARITY.  
 FT DISULFID 515 528 BY SIMILARITY.  
 FT DISULFID 531 540 BY SIMILARITY.  
 FT DISULFID 544 560 BY SIMILARITY.

FT DISULFID 563 576 BY SIMILARITY.  
 FT DISULFID 567 584 BY SIMILARITY.  
 FT DISULFID 587 596 BY SIMILARITY.  
 FT DISULFID 600 623 BY SIMILARITY.  
 FT DISULFID 626 634 BY SIMILARITY.  
 FT DISULFID 630 642 BY SIMILARITY.  
 FT MOD\_RES 1139 1139 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT MOD\_RES 1248 1248 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 124 124 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 187 187 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 259 259 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 571 571 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 629 629 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT VARIANT 654 654 I -> V.  
 FT VARIANT 654 654 /FTID-VAR\_004077.  
 FT VARIANT 655 655 I -> V.  
 FT VARIANT 655 655 /FTID-VAR\_004078.  
 FT CONFLICT 1170 1170 P -> A (IN REF. 2).  
 FT SEQUENCE 1255 AA; 137909 MM; 39E9DFDA04DCF962 CRC64;  
 SQ  
 Query Match 95.5%; Score 3776; DB 1; Length 1255;  
 Best Local Similarity 67.9%; Pred. No. 6.9e-280; Indels 336; Gaps 1;  
 Matches 712; Conservative 0; Mismatches 0;  
 QY 1 MELALACRMGLLALLPPGAASIVGCTGTDMLRLPASPETHLDMRLHLYOGCQVVGNL 60  
 DB 1 MELALACRMGLLALLPPGAASIVGCTGTDMLRLPASPETHLDMRLHLYOGCQVVGNL 60  
 QY 61 ELTYLPINASTLSFLODIOEVGVYLIHNOVROYPLORLIRVSTQLPEDNYALAVLDNG 120  
 DB 61 ELTYLPINASTLSFLODIOEVGVYLIHNOVROYPLORLIRVSTQLPEDNYALAVLDNG 120  
 QY 121 DPLNNTPTVPGASGRLRLQRLSTELKGVLIQNNPOLCYODTILMDIFKNNOLA 180  
 DB 121 DPLNNTPTVPGASGRLRLQRLSTELKGVLIQNNPOLCYODTILMDIFKNNOLA 180  
 QY 121 DPLNNTPTVPGASGRLRLQRLSTELKGVLIQNNPOLCYODTILMDIFKNNOLA 180  
 DB 121 DPLNNTPTVPGASGRLRLQRLSTELKGVLIQNNPOLCYODTILMDIFKNNOLA 180  
 QY 181 LTLIDTNRSRACHPCSPMGKSRGKSSSEDCQSLRTTVCAAGCARCKGLPDDCCEOC 240  
 DB 181 LTLIDTNRSRACHPCSPMGKSRGKSSSEDCQSLRTTVCAAGCARCKGLPDDCCEOC 240  
 QY 241 AAGCTGPKHSDCLALCFHNSHIGELHCPALVYNTDTFESMPNDEGRYTFGASCVTAP 300  
 DB 241 AAGCTGPKHSDCLALCFHNSHIGELHCPALVYNTDTFESMPNDEGRYTFGASCVTAP 300  
 QY 241 AAGCTGPKHSDCLALCFHNSHIGELHCPALVYNTDTFESMPNDEGRYTFGASCVTAP 300  
 DB 241 AAGCTGPKHSDCLALCFHNSHIGELHCPALVYNTDTFESMPNDEGRYTFGASCVTAP 300  
 QY 301 YNYLSTVGSCTLVCPPLHNOEVAEDGTORCEKSKPCARVCYGLGMEHLREVAIVTSAN 360  
 DB 301 YNYLSTVGSCTLVCPPLHNOEVAEDGTORCEKSKPCARVCYGLGMEHLREVAIVTSAN 360  
 QY 361 IOEPAGCKITFGSLAFIPESFGDDPASNTAPLAPEOQVFEETLEITGYLYISAMPDLP 420  
 DB 361 IOEPAGCKITFGSLAFIPESFGDDPASNTAPLAPEOQVFEETLEITGYLYISAMPDLP 420  
 QY 421 DLSYFQMLQYRGRILHNGAVSLTLQGLGISWLGSLRLSRLSGLALIHNTHLCEVHTV 480  
 DB 421 DLSYFQMLQYRGRILHNGAVSLTLQGLGISWLGSLRLSRLSGLALIHNTHLCEVHTV 480  
 QY 481 PMQDLFNPQDALLHTANRDEDECVGSLACHQACAGHCHWGPPIQVCNCSQFLRQEC 540  
 DB 481 PMQDLFNPQDALLHTANRDEDECVGSLACHQACAGHCHWGPPIQVCNCSQFLRQEC 540  
 QY 541 VEECRVLOGLPREVYVNAHCLPCHPECPONGSVTCGPADOCVACAHKDDPPFCVAC 600  
 DB 541 VEECRVLOGLPREVYVNAHCLPCHPECPONGSVTCGPADOCVACAHKDDPPFCVAC 600  
 QY 601 PSQVKKPLSLYMPIMKFPDEGACQPCINCTHSCVDIDDKGCPAEGASPLTS----- 653  
 DB 601 PSQVKKPLSLYMPIMKFPDEGACQPCINCTHSCVDIDDKGCPAEGASPLTS----- 653  
 QY 654 ----- 653  
 DB 661 ILVVVLGVVFGILIKRROOKIRKTYMRRLLOETELVEPLPSGAMENQAMRLKETEL 720

QY	654	-----	653	CC	-1- SUBCELLULAR LOCATION: Type I membrane protein.
CC			CC	CC	-1- PTR: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC			CC	CC	RESIDUES (BY SIMILARITY).
CC			CC	CC	-1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC			CC	CC	-----
CC			CC	CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC			CC	CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC			CC	CC	the European Bioinformatics Institute. There are no restrictions on its
CC			CC	CC	use by non-profit institutions as long as its content is in no way
CC			CC	CC	modified and this statement is not removed. Usage by and for commercial
CC			CC	CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>
CC			CC	CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
CC			CC	CC	-----
DR	EMBL; X03362; CAA27059.1; ALT_INIT.		DR	EMBL; X03362; CAA27059.1; ALT_INIT.	
DR	PIR; A24362; TVRTNU.		DR	PIR; A24362; TVRTNU.	
DR	HSSP; P11362; IFKG.		DR	HSSP; P11362; IFKG.	
DR	InterPro: IPR000494; EGFR_L_domain.		DR	InterPro: IPR000494; EGFR_L_domain.	
DR	InterPro: IPR000719; Euk_kinase.		DR	InterPro: IPR000719; Euk_kinase.	
DR	InterPro: IPR002174; Furin-like.		DR	InterPro: IPR002174; Furin-like.	
DR	InterPro: IPR001245; Tyr_kinase.		DR	InterPro: IPR001245; Tyr_kinase.	
DR	InterPro: IPR004019; YLP_motif.		DR	InterPro: IPR004019; YLP_motif.	
DR	Pfam; PF00069; Kinase; 1.		DR	Pfam; PF00069; Kinase; 1.	
DR	Pfam; PF00757; Furin-like; 1.		DR	Pfam; PF00757; Furin-like; 1.	
DR	Pfam; PF01030; Recep_L_domain; 2.		DR	Pfam; PF01030; Recep_L_domain; 2.	
DR	Pfam; PF02757; YLP; 2.		DR	Pfam; PF02757; YLP; 2.	
DR	ProDom; PD000001; Euk_kinase; 1.		DR	ProDom; PD000001; Euk_kinase; 1.	
DR	SMART; SM00261; FU; 3.		DR	SMART; SM00261; FU; 3.	
DR	SMART; SM00219; Tyrc; 1.		DR	SMART; SM00219; Tyrc; 1.	
DR	PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.		DR	PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.	
DR	PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.		DR	PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.	
DR	PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.		DR	PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.	
KW	Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;		KW	Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;	
KW	transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;		KW	transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;	
KW	Pico-Oncogene; Disease mutation.		KW	Pico-Oncogene; Disease mutation.	
FT	SIGNAL	1	FT	SIGNAL	21
FT	CHAIN	22	FT	CHAIN	1257
FT	DOMAIN	22	FT	DOMAIN	654
FT	TRANSMEM	655	FT	TRANSMEM	677
FT	DOMAIN	678	FT	DOMAIN	1257
FT	DOMAIN	159	FT	DOMAIN	369
FT	DOMAIN	473	FT	DOMAIN	646
FT	DOMAIN	722	FT	DOMAIN	989
FT	NP_BIND	728	FT	NP_BIND	736
FT	BINDING	755	FT	BINDING	755
FT	ACT_SITE	847	FT	ACT_SITE	847
FT	DISULFID	196	FT	DISULFID	205
FT	DISULFID	196	FT	DISULFID	205
FT	DISULFID	200	FT	DISULFID	213
FT	DISULFID	221	FT	DISULFID	228
FT	DISULFID	225	FT	DISULFID	236
FT	DISULFID	237	FT	DISULFID	245
FT	DISULFID	241	FT	DISULFID	253
FT	DISULFID	256	FT	DISULFID	265
FT	DISULFID	269	FT	DISULFID	296
FT	DISULFID	300	FT	DISULFID	312
FT	DISULFID	316	FT	DISULFID	332
FT	DISULFID	335	FT	DISULFID	339
FT	DISULFID	513	FT	DISULFID	522
FT	DISULFID	517	FT	DISULFID	530
FT	DISULFID	533	FT	DISULFID	542
FT	DISULFID	546	FT	DISULFID	562
FT	DISULFID	565	FT	DISULFID	578
FT	DISULFID	569	FT	DISULFID	586
FT	DISULFID	589	FT	DISULFID	598
FT	DISULFID	602	FT	DISULFID	625
FT	DISULFID	628	FT	DISULFID	636
FT	DISULFID	632	FT	DISULFID	644
FT	MOD_RES	1141	FT	MOD_RES	1141
FT	MOD_RES	1250	FT	MOD_RES	1250
FT	CARBOHYD	68	FT	CARBOHYD	68
FT	CARBOHYD	188	FT	CARBOHYD	188
FT	CARBOHYD	260	FT	CARBOHYD	260
FT	CARBOHYD	532	FT	CARBOHYD	532
FT	CARBOHYD	573	FT	CARBOHYD	573

RESULT 2

ERB2\_RAT

ID ERB2\_RAT STANDARD: PRT; 1257 AA.

AC P06494;

DT 01-JAN-1988 (Rel. 06, Created)

DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)

DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Epidermal growth factor

DE receptor-related protein).

GN ERBB2 OR NEU.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI\_TaxID=10116;

RN [1]

RP \*SEQUENCE FROM N.A.

RP TISSUE=Neuroblastoma;

RX \*MEDLINE=66118662; PubMed=3945311;

RA Bargmann C.I., Hung M.-C., Weinberg R.A.;

RT "The neu oncogene encodes an epidermal growth factor receptor-related

RT protein.";

RT Nature 319:226-230(1986).

RN [2]

RP SEQUENCE OF 852-905 FROM N.A.

RC TISSUE=Sciatic nerve;

RX \*MEDLINE=91222560; PubMed=2025425;

RA Lal C., Lemke G.;

RT "An extended family of protein-tyrosine kinase genes differentially

RT expressed in the vertebrate nervous system.";

RT Neuron 6:691-704(1991).

RL [3]

RN STRUCTURE BY NMR OF 650-668.

RP MEDLINE=92155181; PubMed=1346763;

RA Guillek W.J., Bottomley A.C., Iofis F.J., Doak D.G., Mulvey D.,

RA Newman R., Crumpton M.J., Sternberg M.J.E., Campbell I.D.;

RT "Three dimensional structure of the transmembrane region of the proto-

RT oncogenic and oncogene forms of the neu protein.";

RT EMBO J. 11:43-48(1992).

CC -1- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,

CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A

CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-

CC ALPHA AND AMPHIREGULIN.

CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine - ADP + protein

CC tyrosine phosphate.

CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS.

CC THE CONSTITUTIVELY ACTIVATED ONCOGENIC VARIANT FORMS A HOMODIMER.

FT CARBOHYD 631 631 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT VARIANT 661 661 V -> E (IN ONCOGENIC NEU).  
 SQ SEQUENCE 1257 AA; 138831 MW; 6129264583011402 CRC64;

Query Match 81.1%; Score 3205; DB 1; Length 1257;  
 Best Local Similarity 57.9%; Pred. No. 2.3e-236;  
 Matches 608; Conservative 34; Mismatches 70; Indels 338; Gaps 3;

QY 1 METALCRMGILLALLPFGASTVCTGTMKRLPASPETHLDMLRHLVQCGVVOGNTL 60  
 DB 1 METALCRMGILLALLPFGASTVCTGTMKRLPASPETHLDMLRHLVQCGVVOGNTL 60  
 QY 61 ELTYLPNTASLSFLQDIQEVGYVLAHNOVROVPLQRLIVRGTOLEEDNVALAVLNDG 120  
 DB 61 ELTYLPNTASLSFLQDIQEVGYVLAHNOVROVPLQRLIVRGTOLEEDNVALAVLNDG 120  
 QY 121 DPLNTTPVT-GASPGRLQRLSLTEILKGVLTQRNQLCYODIILMKDIFHNKQL 179  
 DB 121 DPLNTTPVT-GASPGRLQRLSLTEILKGVLTQRNQLCYODIILMKDIFHNKQL 179  
 QY 121 DPQDNAASTPGRPGLRQLRLSLTEILKGVLTQRNQLCYODIILMKDIFHNKQL 180  
 DB 121 DPQDNAASTPGRPGLRQLRLSLTEILKGVLTQRNQLCYODIILMKDIFHNKQL 180  
 QY 180 ATLIDITNRSRACHPCSPMCKGSRGSESEDCQSLTRFYCAGGACARCKGPLPTDCHEQ 239  
 DB 180 ATLIDITNRSRACHPCSPMCKGSRGSESEDCQSLTRFYCAGGACARCKGPLPTDCHEQ 239  
 QY 181 APVDITNRSRACHPCSPMCKGSRGSESEDCQSLTRFYCAGGACARCKGPLPTDCHEQ 240  
 DB 181 APVDITNRSRACHPCSPMCKGSRGSESEDCQSLTRFYCAGGACARCKGPLPTDCHEQ 240  
 QY 240 CAAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDFESMPNEGRYTFGASCVTAC 299  
 DB 240 CAAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDFESMPNEGRYTFGASCVTAC 299  
 QY 300 PNYVLSTDVSSCTLVCPRLHNOEYTAEDGTCRCKSPCARVCGYGLMEHLREVRATSA 359  
 DB 300 PNYVLSTDVSSCTLVCPRLHNOEYTAEDGTCRCKSPCARVCGYGLMEHLREVRATSA 359  
 QY 301 PNYVLSTDVSSCTLVCPRLHNOEYTAEDGTCRCKSPCARVCGYGLMEHLREVRATSA 360  
 DB 301 PNYVLSTDVSSCTLVCPRLHNOEYTAEDGTCRCKSPCARVCGYGLMEHLREVRATSA 360  
 QY 360 NIOEFGCKKIFGSLAFLPSPFGDDPASNAPLOPEQLQVFEETLEITGYLYISAMPDSL 419  
 DB 360 NIOEFGCKKIFGSLAFLPSPFGDDPASNAPLOPEQLQVFEETLEITGYLYISAMPDSL 419  
 QY 361 NIOEFGCKKIFGSLAFLPSPFGDDPASNAPLOPEQLQVFEETLEITGYLYISAMPDSL 420  
 DB 361 NIOEFGCKKIFGSLAFLPSPFGDDPASNAPLOPEQLQVFEETLEITGYLYISAMPDSL 420  
 QY 420 PDLSEFONLOVIRGLIHANGAYSLTLOGICISWGLRSLRELSGALLHNNHLCFVHT 479  
 DB 420 PDLSEFONLOVIRGLIHANGAYSLTLOGICISWGLRSLRELSGALLHNNHLCFVHT 479  
 QY 421 RDLSEFONLOVIRGLIHANGAYSLTLOGICISWGLRSLRELSGALLHNNHLCFVHT 480  
 DB 421 RDLSEFONLOVIRGLIHANGAYSLTLOGICISWGLRSLRELSGALLHNNHLCFVHT 480  
 QY 480 VPMDOLFRNHQALLHNSGNPEEDLVSSGLVCSLCAHGHCMGPGTQCVNCSOPIRGO 538  
 DB 480 VPMDOLFRNHQALLHNSGNPEEDLVSSGLVCSLCAHGHCMGPGTQCVNCSOPIRGO 538  
 QY 481 VPMDOLFRNHQALLHNSGNPEEDLVSSGLVCSLCAHGHCMGPGTQCVNCSOPIRGO 540  
 DB 481 VPMDOLFRNHQALLHNSGNPEEDLVSSGLVCSLCAHGHCMGPGTQCVNCSOPIRGO 540  
 QY 539 ECVEECFVLOGLPREVYNAHCLPCHPECOPONSVTCFPEADQCYACAHYKDPFCA 598  
 DB 539 ECVEECFVLOGLPREVYNAHCLPCHPECOPONSVTCFPEADQCYACAHYKDPFCA 598  
 QY 541 ECVEECFVLOGLPREVYNAHCLPCHPECOPONSVTCFPEADQCYACAHYKDPFCA 600  
 DB 541 ECVEECFVLOGLPREVYNAHCLPCHPECOPONSVTCFPEADQCYACAHYKDPFCA 600  
 QY 599 RCFSGVAPDLSTYPMTKFPEBEACQCPINCHSCVDLDDKGCRAQORASPLT----- 652  
 DB 599 RCFSGVAPDLSTYPMTKFPEBEACQCPINCHSCVDLDDKGCRAQORASPLT----- 652  
 QY 601 RCFSGVAPDLSTYPMTKFPEBEACQCPINCHSCVDLDDKGCRAQORASPLT----- 660  
 DB 601 RCFSGVAPDLSTYPMTKFPEBEACQCPINCHSCVDLDDKGCRAQORASPLT----- 660  
 QY 653 ----- 652  
 DB 653 ----- 652  
 QY 661 VGVLFLILVVVVGILIKRRKRIKRYTMRLLOETELVEPLTPSGAMPQAOIRILKET 720  
 DB 661 VGVLFLILVVVVGILIKRRKRIKRYTMRLLOETELVEPLTPSGAMPQAOIRILKET 720  
 QY 653 ----- 652  
 DB 653 ----- 652  
 QY 721 ELRKVVILSGAFGVYKGIWIPDGENVKIPAIKVLRENTSPRANKIILDEAYVAGVG 780  
 DB 721 ELRKVVILSGAFGVYKGIWIPDGENVKIPAIKVLRENTSPRANKIILDEAYVAGVG 780  
 QY 653 ----- 652  
 DB 653 ----- 652  
 QY 781 SPYVSRLLIGLICLTSTVOLVTLQMPYGLLDHVRHGRGLSGODLLMNCVOIANKMSYLE 840  
 DB 781 SPYVSRLLIGLICLTSTVOLVTLQMPYGLLDHVRHGRGLSGODLLMNCVOIANKMSYLE 840  
 QY 653 ----- 652  
 DB 653 ----- 652  
 QY 841 VRLVHRLAARNVLYKSPNHYKITDFGLARLLIDETEHADGSKVPIKMALESILRR 900  
 DB 841 VRLVHRLAARNVLYKSPNHYKITDFGLARLLIDETEHADGSKVPIKMALESILRR 900  
 QY 653 ----- 652  
 DB 653 ----- 652  
 QY 901 FTHQSDVWGYVWELMTGAKPYDGIAPAREIPDLLEKGRRLPQPICTIDVYINWKC 960  
 DB 901 FTHQSDVWGYVWELMTGAKPYDGIAPAREIPDLLEKGRRLPQPICTIDVYINWKC 960  
 QY 653 ----- 652  
 DB 653 ----- 652

DB 961 WVIDSECRPRFRELVEFSFMRADPQRFVYIQNEDLGSSPMSTYSRLLEDDMDGLV 1020  
 QY 683 DAEYLVPOOGFCPPDPAPGAGGVHHRH 712  
 DB 1021 DAEYLVPOOGFCPPDPAPGAGGVHHRH 1050

RESULT 3  
 ERB2\_MESAU  
 ID ERB2\_MESAU STANDARD; PRT; 1254 AA.  
 AC 060553;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)  
 DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2).  
 GN ERB2 OR NEU.  
 OS Mesocricetus auratus (Golden hamster).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
 OC Mesocricetus.  
 OX NCBI\_TaxId=10036;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Nerve;  
 RX MEDLINE=94193007; PubMed=7908275;  
 RA Nakamura T., Ushijima T., Ishizaka Y., Nagao M., Arai M.,  
 RA Yamazaki Y., Ishikawa T.;  
 RT Cloning and activation of the Syrian hamster neu proto-oncogene.";  
 RL Gene 140:251-255(1994).  
 CC -1- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,  
 CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A  
 CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-  
 CC ALPHA AND AMPHIREGULIN (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 CC tyrosine phosphate.  
 CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS  
 CC (POTENTIAL).  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE  
 CC RESIDUES.  
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; D16295; BAA03801.1; -  
 CC HSSP; P11362; 1FGK.  
 DR InterPro: IPR000494; EGFR\_L.domain.  
 DR InterPro: IPR000719; Euk\_Pkinase.  
 DR InterPro: IPR002174; Furin-like.  
 DR InterPro: IPR001245; Tyr\_Pkinase.  
 DR InterPro: IPR004019; YLP\_motif.  
 DR Pfam: PF00069; Pkinase; 1.  
 DR Pfam: PF00757; Furin-like; 1.  
 DR Pfam: PF01030; Recep\_L.domain; 2.  
 DR Pfam: PF02757; YLP; 2.  
 DR ProDom: PD000001; Euk\_Pkinase; 1.  
 DR SMART: SM00219; TyrKc; 1.  
 DR SMART: PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR; 1.  
 DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;  
 KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;  
 KW Proto-oncogene; Disease mutation.  
 FT SIGNAL 1 21 POTENTIAL.

```
FT CHAIN 22 1254 RECEPTOR PROTEIN-TYROSINE KINASE ERB-B-2.
FT DOMAIN 22 652 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 653 675 POTENTIAL.
FT DOMAIN 676 1254 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 158 368 CYS-RICH.
FT DOMAIN 472 644 CYS-RICH.
FT DOMAIN 720 987 PROTEIN KINASE.
FT NP_BIND 726 734 ATP (BY SIMILARITY).
FT BINDING 753 753 ATP (BY SIMILARITY).
FT ACT_SITE 845 845 BY SIMILARITY.
FT DISULFID 195 204 BY SIMILARITY.
FT DISULFID 199 212 BY SIMILARITY.
FT DISULFID 236 244 BY SIMILARITY.
FT DISULFID 240 252 BY SIMILARITY.
FT DISULFID 255 264 BY SIMILARITY.
FT DISULFID 268 295 BY SIMILARITY.
FT DISULFID 299 311 BY SIMILARITY.
FT DISULFID 315 331 BY SIMILARITY.
FT DISULFID 334 338 BY SIMILARITY.
FT DISULFID 511 520 BY SIMILARITY.
FT DISULFID 515 528 BY SIMILARITY.
FT DISULFID 531 540 BY SIMILARITY.
FT DISULFID 544 560 BY SIMILARITY.
FT DISULFID 563 576 BY SIMILARITY.
FT DISULFID 567 584 BY SIMILARITY.
FT DISULFID 587 596 BY SIMILARITY.
FT DISULFID 600 623 BY SIMILARITY.
FT DISULFID 626 634 BY SIMILARITY.
FT MOD_RES 1139 642 BY SIMILARITY.
FT MOD_RES 1247 1247 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 68 68 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 125 125 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 187 187 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 259 259 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 530 530 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 571 571 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 629 629 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 658 658 V -> E (IN ONCOGENIC NEU).
FT VARIANT 659 659 V -> E (IN ONCOGENIC NEU).
SQ SEQUENCE 1254 AA; 138252 MW; 974C3791C21F2BE1 CRC64;

Query Match 80.6%; Score 3185; DB 1; Length 1254;
Best Local Similarity 57.5%; Pred. No. 7.5e-235;
Matches 603; Conservative 36; Mismatches 73; Indels 336; Gaps 1;
```

```
OY 421 DLSVFQNIQVIRGRITLHNGAVSLTQIGISWLGRLREIGSSGLALIHNNHLCFVHTV 480
DB 421 DLSVFQNIQVIRGRITLHNGAVSLTQIGISWLGRLREIGSSGLALIHNNHLCFVHTV 480
OY 481 PMDQLFRRNPQALHTANRPDECEGEGELACQALCARGCMGPGPTOCVNCSPFRGEC 540
DB 481 PMDQLFRRNPQALHTANRPDECEGEGELACQALCARGCMGPGPTOCVNCSPFRGEC 540
OY 541 VEECRVYLGPREVYNARHCLPCHECOPQNGSVTCFGEADQVACAHYKDPPECVARC 600
DB 541 VEECRVYLGPREVYNARHCLPCHECOPQNGSVTCFGEADQVACAHYKDPPECVARC 600
OY 541 VEECRVYLGPREVYNARHCLPCHECOPQNGSVTCFGEADQVACAHYKDPPECVARC 600
DB 541 VEECRVYLGPREVYNARHCLPCHECOPQNGSVTCFGEADQVACAHYKDPPECVARC 600
OY 601 PSQVPRDLSYMHYKPFPEDEGACQPCPTNCCHSCVDLDDKGPAPQARSPLTS 653
DB 601 PSQVPRDLSYMHYKPFPEDEGACQPCPTNCCHSCVDLDDKGPAPQARSPLTS 653
OY 601 PSQVPRDLSYMHYKPFPEDEGACQPCPTNCCHSCVDLDDKGPAPQARSPLTS 653
DB 601 PSQVPRDLSYMHYKPFPEDEGACQPCPTNCCHSCVDLDDKGPAPQARSPLTS 653
OY 654 ----- 653
DB 661 ILFLVIGVVGILIKRRRQKIRKTYMRRLQETELVEPLTPSGAMPNOQMRLKETEL 720
OY 654 ----- 653
DB 721 RKVKVIGSGAGCTVYKGIWIPDGENVKIPAIKVLRENTSPRANKELIDEAYVAGLSP 780
OY 654 ----- 653
DB 781 YVSRLLGICLSTVOLVTLQMPYGLDLHVRHGRGLSGDILNMCVOIAKMSYLEDVR 840
OY 654 ----- 653
DB 841 LVHRLAARNVYKSPNHVKTIDFGLARLIDIDEYHADGKVIKMALESILRRFT 900
OY 654 ----- 653
DB 901 HQSDVSYGVYVWELMTFGAKPYDGIIPAREIPDLLEKGBRLQPPICITIDVYMIWKCMM 960
OY 654 ----- 684
DB 961 IDSECRPRFRELVSFSESRMARDPQRFVYIOMEDLSPSSPLSTFYRSLLEDMDGLDVA 1020
OY 685 EEYLVPOGFCPPDPAPGAGGVHHRHR 712
DB 1021 EEYLVPOGFCPPDPAPGAGGVHHRHR 1048

RESULT 4
EGFR_MOUSE STANDARD: PRT: 1210 AA.
AC 001279;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112).
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Balb/c; TISSUE=Liver;
RX MEDLINE=93026370; PubMed=1408137;
RA Avivi A., Skorecki K., Yaron A., Givol D.;
RT "Promoter region of the murine fibroblast growth factor receptor 2
RT (hek/KGFR) gene."
RT Oncogene 7:1957-1962(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Balb/c, and CD-1; TISSUE=Liver, and Decidua;
RX MEDLINE=93126380; PubMed=7678348;
RA Paria B.C., Das S.K., Andrews G.K., Dey S.K.;
RT "Expression of the epidermal growth factor receptor gene is regulated
```



QY 244 CTGPKHSDCLACLNHNSHSGICELHCPALVYNTDTEESMPNDEGRYTGASCVTACAPYNY 303  
 DB 240 CTGPRSDPCLVQKQKQDQATCKDCTPPLMLYNPTTYQMDVNBEGKYSRGATCKKCPKNY 299  
 QY 304 LSTDVSGCTVLCPLHNOVETADGTQRCCKSPCARVCYGLGMEHLREAVRASANTIOE 363  
 DB 300 VYTDHSGSCVRACGPDYEV-EDDGIRCKCKDGCGRKVCNNGIGIEFDYLTINATNIKH 358  
 QY 364 FAGCKKIGSLAFLESPSGDPSASTAPLOEQLOVEFTLEETGYLXIAMPDLSPLDS 423  
 DB 359 FKYCAISGDLHLPVAKGDSFTPTPLDRELEITLVEETGELLIQAMPDMMWTLH 418  
 QY 424 VFONLQVIRGTHLHNGAVSLTLOGIGISMLRLSRLRELSGALIHHTHLCFVETVPM 483  
 DB 419 AFENELIRGRTRKQHGQSLAVNGINTSLGRSLKETISDDVITISGNRLCYANTIMWK 478  
 QY 484 QLFNPHOALLHTANRPDECEVGEGLACHQLCARGHCGPPTQCVNSQPLRGCEVVEE 543  
 DB 479 KLFETPNQTKIMNNRAEKDKAVNHVGNPLCSSGCGMPRPDVCQNTSRGREGVEK 538  
 QY 544 CRVIGLQPREVYNARHCLPCHPEQOPONGSVTCGPREADQVACAHYNDPPFCVARGPSG 603  
 DB 539 CNLIEGPERVEVSECTIQCHPECLPOMANTCTGRGPDNCTQCAHYIDGPHCVKTCBAG 598  
 QY 604 VKPDLSTYMPKPFDEEGACQPCPINCTHSCVDLDDKGC 642  
 DB 599 IMGENTNL-VWKYADANNVCHLCHANCYGCAGPLQGC 636

## RESULT 5

ID EGFR\_HUMAN STANDARD: PRT: 1210 AA.

AC P00533; P06268; Q14225; Q9UMD7; Q9UMD8; Q9UMG5; Q92795; 000732;  
 AC Q00688; Q9B852; Q9H2C9; Q9G2X1; Q9H3C9;  
 DT 21-JUL-1986 (Rel. 01, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Receptor

protein-tyrosine kinase EribB-1).

GN EGFR OR EKBBI.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1);

RX MEDLINE-84219729; PubMed-6328312;  
 RA Ullrich A., Coussens L., Hayflick J.S., Dull T.J., Gray A., Tam A.W.,  
 RA Lee J., Yarden Y., Libermann T.A., Schlessinger J., Downward J.,  
 RA Mayes E.L.V., Whittle N., Waterfield M.D., Seeburg P.H.;

RT "Human epidermal growth factor receptor cDNA sequence and aberrant  
 RT expression of the amplified gene in A431 epidermoid carcinoma cells."

RL Nature 309:418-425(1984).

RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).

RX MEDLINE-95382957; PubMed-7654368;  
 RA Ilekis J.V., Stark B.C., Scoccia B.;

RT "Possible role of variant RNA transcripts in the regulation of  
 RT epidermal growth factor receptor expression in human placenta.";

RL Mol. Reprod. Dev. 41:149-156(1995).

RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).

RX MEDLINE-97078686; PubMed-8918811;  
 RA Reller J.L., Maible N.J.;

RT "A 1.8 kb alternative transcript from the human epidermal growth  
 RT factor receptor gene encodes a truncated form of the receptor.";

RL Nucleic Acids Res. 24:4050-4056(1996).  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RC TISSUE-Placenta;  
 RX MEDLINE-97256547; PubMed-9103388;

RA Ilekis J.V., Gariti J., Niederberger C., Scoccia B.;

RT "Expression of a truncated epidermal growth factor receptor-like  
 RT protein (TEGFR) in ovarian cancer.";

RL Gynecol. Oncol. 65:36-41(1997).

RN [5]  
 RP SEQUENCE FROM N.A. (ISOFORMS 3 AND 4).

RC TISSUE-Placenta;  
 RX MEDLINE-21100872; PubMed-11161793;

RA Reller J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,  
 RA Schehl Sinclair C., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,  
 RA Balasubramaniam S., Crossley T.D., Magnuson T.R., James C.D.;

RT "Comparative genomic sequence analysis and isolation of human and  
 RT mouse alternative EGFR transcripts encoding truncated receptor  
 RT isoforms.";

RL Genomics 71:1-20(2001).

RN [6]  
 RP SEQUENCE OF 575-687 FROM N.A.

RA Reller J.L., Threadgill D.W., Danielson A.J., Schehl C.M.,  
 RA Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,  
 RA Maible N.J.;

RT "Human and mouse alternative EGFR transcripts encoding only the  
 RT extracellular domain of the receptor.";

RL Submitted (FEB-1999) to the EMBL/Genbank/DBJ databases.

RN [7]  
 RP SEQUENCE OF 713-924 FROM N.A.

RX MEDLINE-84196372; PubMed-6326261;  
 RA Lin C.R., Chen W.S., Kruliger W., Stolarsky L.S., Weber W.,  
 RA Evans R.M., Verma I.M., Gill G.N., Rosenfield M.G.;

RT "Expression cloning of human EGF receptor complementary DNA: gene  
 RT amplification and three related messenger RNA products in A431  
 RT cells.";

RL Science 224:843-848(1984).

RN [8]  
 RP SEQUENCE OF 150-962 FROM N.A.

RX MEDLINE-84245835; PubMed-6330563;  
 RA Xu Y.H., Ishii S., Clark A.J.L., Sullivan M., Wilson R.K., Ma D.P.,  
 RA Roe B.A., Merlino G.T., Pastan I.,

RT "Human epidermal growth factor receptor cDNA is homologous to a  
 RT variety of RNAs overproduced in A431 carcinoma cells.";

RL Nature 309:806-810(1984).

RN [9]  
 RP SEQUENCE OF 1028-1210 FROM N.A.

RX MEDLINE-85046483; PubMed-6093780;  
 RA Stimmgen F.A., Gope M.L., Schultz T.Z., Wright D.A., Carpenter G.,  
 RA O'Walley B.W.;

RT "Isolation of an evolutionarily conserved epidermal growth factor  
 RT receptor cDNA from human A431 carcinoma cells.";

RL Biochem. Biophys. Res. Commun. 124:125-132(1984).

RN [10]  
 RP SEQUENCE OF 1-29 FROM N.A.

RX MEDLINE-88217333; PubMed-3329716;  
 RA Haley J.D., Whittle N., Bennett P., Kinchington D., Ullrich A.,  
 RA Waterfield M.D.;

RT "The human EGF receptor gene: structure of the 110 kb locus and  
 RT identification of sequences regulating its transcription.";

RL Oncogene Res. 1:373-396(1987).

RN [11]  
 RP SEQUENCE OF 1-29 FROM N.A.

RX MEDLINE-91107677; PubMed-1988448;  
 RA Haley J.D., Waterfield M.D.;

RT "Contributory effects of de novo transcription and premature  
 RT transcript termination in the regulation of human epidermal growth  
 RT factor receptor proto-oncogene RNA synthesis.";

RL J. Biol. Chem. 266:1746-1753(1991).

RN [12]  
 RP SEQUENCE OF 1-29 FROM N.A.

RX MEDLINE-85270438; PubMed-2991899;  
 RA Ishii S., Xu Y.H., Stratton R.H., Roe B.A., Merlino G.T., Pastan I.;

RT "Characterization and sequence of the promoter region of the human  
 RT epidermal growth factor receptor gene.";

RL Proc. Natl. Acad. Sci. U.S.A. 82:4920-4924(1985).  
 RN [13]

RP SEQUENCE OF 540.  
RA Kohda D.;  
RL Submitted (SEP-1997) to the SWISS-PROT data bank.  
RN [14]  
RP RECEPTOR ACTIVITY.  
RX MEDLINE=84191554; PubMed=6325948;  
RA Mroczkowski B., Mosig G., Cohen S.;  
RT "ATP-stimulated interaction between epidermal growth factor receptor  
and supercoiled DNA".  
RL Nature 309:270-273(1984).  
RN [15]  
RP PHOSPHORYLATION.  
RX MEDLINE=89278137; PubMed=2533678;  
RA Margolis B.L., Iax I., Kris R., Dombalagian M., Honegger A.M.,  
RA Howk R., Glyol D., Ullrich A., Schlessinger J.;  
RT "All autophosphorylation sites of epidermal growth factor (EGF)  
receptor and HER2/neu are located in their carboxyl-terminal tails.  
Identification of a novel site in EGF receptor".  
RL J. Biol. Chem. 264:10667-10671(1989).  
RN [16]  
RP CARBOHYDRATE-LINKAGE SITES ASN-128; ASN-175; ASN-413; ASN-444 AND  
ASN-528.  
RX MEDLINE=96398132; PubMed=8962717;  
RA Smith K.D., Davies M.J., Bailey D., Renouf D.V., Hounsell E.F.;  
RT "Analysis of the glycosylation patterns of the extracellular domain of  
the epidermal growth factor receptor expressed in Chinese hamster  
ovary fibroblasts".  
RL Growth Factors 13:121-132(1996).  
RN [17]  
RP CARBOHYDRATE-LINKAGE SITES ASN-56; ASN-352; ASN-361; ASN-568 AND  
ASN-603.  
RX MEDLINE=20198209; PubMed=10731668;  
RA Sato C., Kim J.-H., Abe Y., Saito K., Yokoyama S., Kohda D.;  
RT "Characterization of the N-Oligosaccharides attached to the atypical  
Asn-X-Cys sequence of recombinant human epidermal growth factor  
receptor".  
RL J. Biochem. 127:65-72(2000).  
RN [18]  
RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.  
RX MEDLINE=98225196; PubMed=9556602;  
RA Abe Y., Otake M., Inagaki F., Iax I., Schlessinger J., Kohda D.;  
RT "Disulfide bond structure of human epidermal growth factor receptor".  
RL J. Biol. Chem. 273:11150-11157(1998).  
RN [19]  
RP REVIEW.  
RX MEDLINE=87297456; PubMed=3039909;  
RA Carpenter G.;  
RT "Receptors for epidermal growth factor and other polypeptide  
mitogens".  
RL Annu. Rev. Biochem. 56:881-914(1987).  
CC -1- FUNCTION: Receptor for EGF, but also for other members of the EGF  
family, as TGF-alpha, amphiregulin, betacellulin, heparin-binding  
EGF-like growth factor, GP30 and vaccinia virus growth factor. Is  
involved in the control of cell growth and differentiation.  
CC -1- FUNCTION: Isoform 2/truncated isoform may act as an antagonist.  
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
tyrosine phosphate.  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Isoform 2 is  
secreted.  
CC -1- ALTERNATIVE PRODUCTS: 4 isoforms: 1/p170 (shown here), 2/p60/  
truncated isoform/TEGFR, 3/p110 and 4; are produced by  
alternative splicing.  
CC -1- TISSUE SPECIFICITY: Expressed in placenta. Isoform 2 is also  
expressed in ovarian cancers.  
CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to  
dimerization, internalization of the EGF-receptor complex,  
induction of the tyrosine kinase activity, stimulation of cell DNA  
synthesis, and cell proliferation.  
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: X00588; CA25240.1; -  
DR EMBL: U95089; AAB53063.1; -  
DR EMBL: U48722; AAC50802.1; -  
DR EMBL: U48723; AAC50804.1; -  
DR EMBL: U48724; AAC50796.1; -  
DR EMBL: U48725; AAC50797.1; -  
DR EMBL: U48726; AAC50798.1; -  
Query\_Match 38.7%; Score 1532; DB 1; Length 1210;  
Best Local Similarity 42.9%; Pred. No. 6,6e-109;  
Matches 283; Conservative 108; Mismatches 241; Indels 28; Gaps 6;  
OY 11 LLLALLPPGAA--STOYCTGDMKRLPASPETHLMDRLHYGGCVOYGNLELTLPYN 68  
DB 11 LLLALLPPGAA--STOYCTGDMKRLPASPETHLMDRLHYGGCVOYGNLELTLPYN 68  
DB 14 LLAALCPASRALBEKKCOGTSNKLTLQGLTFEDHFLSLQMFNNCEVGLNLEITVYORN 73  
OY 69 ASLSFLDIDIOYGVYLIANOVROYPLORLIRVGTQLEFEDYVALVDNGDPLNNTTP 128  
DB 74 YDSFLKTIQEVAGYVLIANTVERIPLENLQIRGMNRYENSYALAVLSNTP----- 126  
OY 129 VTGASPGELRELQRLSTLETLKGVLLIQRNPOCLYODTILMKDIFPKNNQALTLIDTNR 188  
DB 127 ---ANKTGELKELPMRLNQLIELHGAVRFSNNPALCNVESIQMDRIDVSDFLSNMSMDFQNH 183  
OY 189 SRACHPCSPMKCKSRCKGSESEDCOSITRTVCAGGCA-RKGPLPDDCEGCAACTGP 247  
DB 184 LGSCOKDPSCPNPGSCWAGEENCOKLTLLICAQCSGRCRGKSPSCCHNCAACTGP 243  
OY 248 KHSDDLACLEPHNNSGICELCPALVTYNTDTFESMPREGRYFEGASCVACPYNLSTD 307  
DB 244 RESDCLVCRKFRDEATCKDPCPLMLYNTPTTYQMDVNPBEKYSFGATCVKCCRNVTYD 303  
OY 308 VGSCTLVCPHANOEVTAEDGTORCEKSKPCARVCYGLGMEHLREVAVTSANIQEFAG 367  
DB 304 HGSCVVRACGADSYEM-EDGVRCCKKCEGPCRCVNCIGIGERKDSLISNATNIKHKNC 362  
OY 368 KTFPGSLAFLPESFDGDPASNTAPLOPQLQVETLEITGYLYXSNPMSLDLSFQON 427  
DB 363 TSIISGDHLIPVAFRSGSFHTPDLQDELIDIKTYVEITGFPLLIQAMPNRRLDAFEN 422  
OY 428 LOYRGRILHNGAYSLTLOGAGISWGLRSRLRELSGLALIHNTHTLCFHTVPMQDLFR 487  
DB 423 LEIIRGRTKQHGOFSLAVSLNLTSLGRLSKELISDGDVLIISGNKNTCYANTIMKKLFG 482  
OY 488 NPHQALLHTANRPDEDCVGEGLACHQACARGHCWPGPPTQCVNCQSLRGQECVEEGRVL 547  
DB 483 TSGGKRTIISNRGNSCKATGVCYCHALCSPEGCGPBPBCVSCRVNSRGRECDKCNLL 542  
OY 548 QGLPREVYNAHCLPCRPCEQOPONGSVTCGFPAPDQCVAAHAKDPFCVAPRPSGKPD 607  
DB 543 EGEPRFEVENSECICQHPDELQAMNITTCGRGPDNDCIOGAHYIDGHCVCPAGWGE 602  
OY 608 LSYWPIKFPDEBAGACPCPINCTHSCVDLDDKGCAPAEGRASPLTSQNEDLGPASPLDST 667  
DB 603 NNTLV-VMKYADAGHVCHLCPNCTYGGCTGPGLEBSCPNT-----GKRIPIAT 648  
RESULT 6  
EGFR\_CHICK STANDARD; PRT; 703 AA.  
AC EGFR\_CHICK P1387;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, last sequence update)  
DT 15-JUN-2002 (Rel. 41, last annotation update)  
DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (CER)  
DE (Fragment).  
GN EGFR.  
OS Gallus gallus (Chicken).





OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM JM-A).  
 RC TISSUE=breast carcinoma;  
 RA MEDLINE=93189574; PubMed=8383326;  
 RA Plozman G.D., Culouscou J.-M., Whitney G.S., Green J.M., Carlton G.W.,  
 RA Foy L., Neuberger M.G., Shoyab M.;  
 RA "Ligand-specific activation of HER4/p180erbB4, a fourth member of the  
 RT epidermal growth factor receptor family";  
 RL Proc. Natl. Acad. Sci. U.S.A. 90:1746-1750(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORMS JM-A AND JM-B).  
 RC TISSUE=Fetal brain;  
 RX MEDLINE=97476287; PubMed=9334263;  
 RA Elenius K., Corfas G., Paul S., Choi C.J., Rio C., Plozman G.D.,  
 RA Klagesbrun M.;  
 RA "A novel juxtamembrane domain isoform of HER4/erbB4, isoform-specific  
 RT tissue distribution and differential processing in response to  
 RT phorbol ester";  
 RL J. Biol. Chem. 272:26761-26768(1997).  
 CC -1- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-  
 CC 2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND  
 CC NTAR. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.  
 CC NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN.  
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 CC tyrosine phosphate.  
 CC -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB  
 CC RECEPTORS (POTENTIAL).  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: JM-A (SHOWN HERE) AND JM-B;  
 CC ARE PRODUCED BY ALTERNATIVE SPLICING. THE TWO ISOFORMS DIFFER  
 CC FUNCTIONALLY IN THEIR RESPONSE TO PHORBOL ESTER. JM-A IS PROCESSED  
 CC BUT NOT JM-B. SO THEY RESPECTIVELY REPRESENT CLEAVABLE AND  
 CC NONCLEAVABLE FORMS OF THE RECEPTOR. BOTH ISOFORMS ARE EXPRESSED IN  
 CC CEREBELLUM, BUT ONLY THE JM-B ISOFORM IS EXPRESSED IN THE HEART.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED AT HIGHEST LEVELS IN BRAIN, HEART,  
 CC KIDNEY, IN ADDITION TO SKELETAL MUSCLE, PARATHYROID, CEREBELLUM,  
 CC PITUITARY, SPLEEN, TESTIS AND BREAST. LOWER LEVELS IN THYMUS,  
 CC LUNG, SALIVARY GLAND, AND PANCREAS.  
 CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE  
 CC RESIDUES.  
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: L07868; AAB59446.1; -;  
 DR HSSP: P11362; 1FGK.  
 DR Genew: HGNC:3432; ERBB4.  
 DR MIM: 600543; -;  
 DR InterPro: IPR000494; EGFR\_L.domain.  
 DR InterPro: IPR000719; Euk\_pkinase.  
 DR InterPro: IPR002174; Furin-like.  
 DR InterPro: IPR001245; Tyr\_pkinase.  
 DR InterPro: IPR004019; YLP\_motif.  
 DR Pfam: PF00069; pkinase.1.  
 DR Pfam: PF00757; Furin-like.1.  
 DR Pfam: PF01030; Recep\_L.domain.2.  
 DR Pfam: PF02757; YLP.2.  
 DR ProDom: PD000001; Euk\_pkinase.1.  
 DR SMART: SM00261; FU.4.  
 DR SMART: SM00219; TYRKc.1.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP.1.  
 DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR.1.  
 DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM.1.  
 KW Transmembrane; Glycoprotein; Multigene family; Receptor; signal;

KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;  
 KW Alternative splicing.  
 FT SIGNAL 1 25  
 FT CHAIN 26 1308  
 FT DOMAIN 26 651  
 FT TRANSMEM 652 675  
 FT DOMAIN 676 1308  
 FT DOMAIN 186 334  
 FT DOMAIN 496 633  
 FT DOMAIN 718 985  
 FT NP\_BIND 724 732  
 FT BINDING 751 751  
 FT ACT\_SITE 843 843  
 FT DISULFID 189 197  
 FT DISULFID 193 205  
 FT DISULFID 213 221  
 FT DISULFID 217 229  
 FT DISULFID 230 238  
 FT DISULFID 234 246  
 FT DISULFID 249 258  
 FT DISULFID 262 289  
 FT DISULFID 293 304  
 FT DISULFID 308 323  
 FT DISULFID 326 330  
 FT DISULFID 303 312  
 FT DISULFID 507 520  
 FT DISULFID 523 532  
 FT DISULFID 536 552  
 FT DISULFID 555 569  
 FT DISULFID 559 577  
 FT DISULFID 580 589  
 FT DISULFID 593 614  
 FT DISULFID 617 625  
 FT DISULFID 621 633  
 FT MOD\_RES 1162 1162  
 FT MOD\_RES 1186 1186  
 FT MOD\_RES 1258 1258  
 FT MOD\_RES 1284 1284  
 FT CARBOHYD 138 138  
 FT CARBOHYD 174 174  
 FT CARBOHYD 181 181  
 FT CARBOHYD 253 253  
 FT CARBOHYD 358 358  
 FT CARBOHYD 410 410  
 FT CARBOHYD 473 473  
 FT CARBOHYD 495 495  
 FT CARBOHYD 548 548  
 FT CARBOHYD 576 576  
 FT CARBOHYD 620 620  
 FT VARSPIC 626 648  
 SQ SEQUENCE 1308 AA; 146807 MW; 5E4A8E095D88761 CRC64;  
 (IN ISOFORM JM-B).  
 Query Match 36.7%; Score 1451; DB 1; Length 1308;  
 Best Local Similarity 43.5%; Pred. No. 1; le-102;  
 Matches 275; Conservatiles 99; Mismatches 238; Indels 20; Gaps 8;  
 QY 9 WGLLALLPGAA-----STVCCTGDMKRLRLPASPERHLMRLHYOGCCVCGVGNLELY 64  
 DB 8 WWWSLVLAAGTYVPPSSQSYCACTENKLSLSLEDOYRALRKRYINCEYVAGNLEITS 67  
 QY 65 LPTNASLFLQDIOEVGYVLIANOVYPLQRLIRVGTQLEFDNYALAVLNDGDP 124  
 DB 68 IEHNRDLSPLRVREYGVVVALNGRYVPLELRLIRRGKLYEDRYALAFINRKDG 127  
 QY 125 NTTPVYTGASPEGLELDLRSLTELKAGVLIQRNPOLCYODTITMKILFIKNNALTLI 164  
 DB 128 NF-----GLQELGLKNTLETINGVYVDKFKFADDTTHMODIVRNPMPSNLTIV 178  
 QY 185 DTNRBACHPCSPMCKGRCGSGESSEDCOSLTRVACAGG-ARCKGGLPTDCHEOCAG 243  
 DB 179 STNSSGCGRCHSCTG-RCKMPTENHCOTILTRIVCAEDCGKRYGVYSDCCRCACAG 237



FT DISULFID 621 633 BY SIMILARITY.  
 FT MOD\_RES 1162 1162 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT MOD\_RES 1188 1188 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT MOD\_RES 1258 1258 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT MOD\_RES 1284 1284 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT CARBOHYD 138 138 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 144 144 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 253 253 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 358 358 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 410 410 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 473 473 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 495 495 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 548 548 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 620 620 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 1062 1062 S -> N (IN REF. 3).  
 FT CONFLICT 1080 1080 PPT -> STR (IN REF. 3).  
 SQ SEQUENCE 1308 AA: 146957 MW: D944B0996A08B41 CRC64;

Query Match 36.4%; Score 1440.5; DB 1; Length 1308;  
 Best Local Similarity 43.7%; Pred. No. 6.8e-102;  
 Matches 279; Conservative 99; Mismatches 242; Indels 19; Gaps 9;

QY 1 MELA-ALCNGLL--ALPFGAASQVCTGDMKRLPASPTFLMDLRLHYGCOVVO 57  
 1 MKLATGLMWGSLIVARVQPSASQVCACTENKLSLSDLOQYRALRKYEEVVM 60  
 QY 58 GNLELYLPFNASTLFDLOEVOGYVLIAHNOVQPLRLRIVRGTOLEFNALAVL 117  
 61 GNLEIISIEHNRDLSTLRSTREYTVYVLANOFRIPLLENKRIKTKIEDRYALATF 120  
 QY 118 DNGDPLNNTPTVYGASPGGLRELOLSLTELKGVLIORNPOLCYODTILMKDIEHKN 177  
 121 LNRKDKGNF-----GLOELGKMLKTEILNGVYVDNCKFLCYADTIHMODIVANPW 171  
 QY 178 QALATLIDITKRSRACHPCSMCKGSRMGSSSDCOSLTFTVACGC-AACKGPIPIPDCC 236  
 172 PSMNTLVTITGSSGCGCHKSCGTG-RCMGPTENHCQTILTFVCAEQDGCRCYGPVSDCC 230  
 QY 237 HECCAGAGCTGPKHSDCLACLPFNHSGICEHCPALVYNTDFESMPNPGRTFGASCV 236  
 231 HRECAGGSSPKPTDPCFACNPNDSGACVQCOTFYNNPTTQLEHNFAKTYGAFV 290  
 QY 297 TACPYNYLSTDVSSCTLVCPDLHNOEYVTAEDGTORCEKSPCARVCYGLMEHLREVRAY 356  
 291 KKCPHNEV-VDSSSCVCAFPSSKMEV-ENGIMKMKRPTDICKACDGIOTGSLMSAQTV 348  
 QY 357 TSANIOEFACKKTFGSLAPESPOGDPASNNAPLOPELOFETLEITGYLISAMP 416  
 349 DSSNIDFEFNGCTINGNLFLVGIHDDPYNAIDALIDPEKLVNFRVREITGFLINTQTP 408  
 QY 417 DSLPDLSEFONLOVIRGRILHNGAYSLTLOGLISWLGSLRELGLALIHNNHLCF 476  
 409 PNNITDSVFENLWITIGSRVLYSGISLILKQOITSLQFSLKELISAGNIYTDNSLKY 468  
 QY 477 VHTVPMDOLFRRNHQALLHNRPEDECVGEGLACHOLCARGHGWGPGPPQVNCOSFLR 536  
 469 YHTINVTTFSTVNORIVINDNRRAENCTAEGVNCNLGNDGCGPGPOCLSCRRFSR 528  
 QY 537 GQSCVEECRYLOGLPREYVARHCLPCHPCOP-ONGSVACFGEADQCAVACAYKRPPE 595  
 529 GKICIESCNLYDGEFFRENGSICVECDSCCKMEDGLTICHGPGPNCTKCSHFKDPN 588  
 DB 589 CVEKCPDYLOGANSF--IFKYADQDRCHCHPCNCTQGC 625

RESULT 9  
 ERR3\_HUMAN  
 ID ERR3\_HUMAN STANDARD; PRT; 1342 AA.  
 AC P21860;  
 DT 01-MAY-1991 (Rel. 18, Created)

DT 01-MAY-1991 (Rel. 18, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)  
 DE (c-erbB3) (Tyrosine kinase-type cell surface receptor HER3).  
 GN ERB3 OR HER3.  
 OS Homo sapiens (Human)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE-90083234; PubMed-2687875;  
 RA Kraus W.H., Issing W., Mikl T., Popescu N.C., Aaronson S.A.;  
 RT "Isolation and characterization of ERB3, a third member of the  
 RT ERB3/epidermal growth factor receptor family: evidence for  
 RT overexpression in a subset of human mammary tumors.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 86:9193-9197(1989).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-90311312; PubMed-2164210;  
 RA Plozman G.D., Whitney G.S., Neubauer M.G., Green J.M., McDonald V.L.,  
 RA Todaro G.J., Shoyab M.;  
 RT "Molecular cloning and expression of an additional epidermal growth  
 RT factor receptor-related gene.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 87:4905-4909(1990).  
 RN [3]  
 RP SEQUENCE FROM N.A. (SHORT FORM).  
 RC TISSUE-Placenta;  
 RX MEDLINE-93282822; PubMed-7685162;  
 RA Katch M., Yazaki Y., Sugimura T., Terada M.;  
 RT "c-erbB3 gene encodes secreted as well as transmembrane receptor  
 RT tyrosine kinase.";  
 RL Biochem. Biophys. Res. Commun. 192:1189-1197(1993).  
 CC 1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTRK.  
 CC 1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 CC tyrosine phosphate.  
 CC 1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERB3 RECEPTORS  
 CC (POTENTIAL).  
 CC 1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG FORM) AND  
 CC SECRETED (SHORT FORM).  
 CC 1- ALTERNATIVE PRODUCTS: TWO FORMS, A LONG TRANSMEMBRANE FORM AND A  
 CC SHORT SECRETED FORM, EXIST DUE TO ALTERNATIVE SPLICING.  
 CC 1- TISSUE SPECIFICITY: EPITHELIAL TISSUES AND BRAIN.  
 CC 1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE  
 CC SH2 OR SH3 DOMAIN OF MANY SIGNAL-TRANSDUCING PROTEINS.  
 CC 1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES  
 CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF  
 CC PHOSPHATIDYLINOSITOL 3-KINASE (BY SIMILARITY).  
 CC 1- DISEASE: OVEREXPRESSED IN A SUBSET OF HUMAN MAMMARY TUMORS.  
 CC 1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sid.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M29366; AAA35790.1; -  
 DR EMBL: M34309; AAA35979.1; -  
 DR EMBL: S61953; AAB26935.1; -  
 DR PIR: A36223; A36223.  
 DR HSSP: P11362; 1FGK.  
 DR Genew; HGNC:3431; ERB3.  
 DR MTM: 190151; -  
 DR InterPro: IPR000494; EGFR\_L\_domain.  
 DR InterPro: IPR000719; Euk\_pkinase.  
 DR InterPro: IPR002174; Furin-like.  
 DR InterPro: IPR001245; Tyr\_pkinase.  
 DR Pfam: PF00069; pkinase; 1.  
 DR Pfam: PF00757; Furin-like; 1.  
 DR Pfam: PF01030; Recep\_L\_domain; 2.

DR ProDom; PD000001; Euk\_pkinase; 1.  
 DR SMART: SM00261; FU; 3.  
 DR SMART: SM00219; TYRKC; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; FALSE\_NEG.  
 DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; FALSE\_NEG.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;  
 KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;  
 KW Alternative splicing.  
 KW SIGNAL  
 FT CHAIN 1  
 FT DOMAIN 20 1342  
 FT TRANSHEM 644 664  
 FT DOMAIN 665 1342  
 FT NP\_BIND 709 966  
 FT BINDING 715 723  
 FT BINDING 742 742  
 FT ACET\_SITE 834 834  
 FT DISULFID 186 194  
 FT DISULFID 190 202  
 FT DISULFID 210 218  
 FT DISULFID 227 235  
 FT DISULFID 231 243  
 FT DISULFID 246 255  
 FT DISULFID 259 286  
 FT DISULFID 290 301  
 FT DISULFID 305 320  
 FT DISULFID 323 327  
 FT DISULFID 500 509  
 FT DISULFID 504 517  
 FT DISULFID 520 529  
 FT DISULFID 533 549  
 FT DISULFID 552 565  
 FT DISULFID 556 573  
 FT DISULFID 576 585  
 FT DISULFID 589 610  
 FT DISULFID 613 621  
 FT DISULFID 617 629  
 FT CARBOHYD 126 126  
 FT CARBOHYD 250 250  
 FT CARBOHYD 353 353  
 FT CARBOHYD 408 408  
 FT CARBOHYD 414 414  
 FT CARBOHYD 437 437  
 FT CARBOHYD 469 469  
 FT CARBOHYD 522 522  
 FT CARBOHYD 566 566  
 FT CARBOHYD 616 616  
 FT VARSPLIC. 141 183  
 FT VARSPLIC. 184 1342  
 FT CONFLICT 560 560  
 FT CONFLICT 1064 1064  
 FT SEQUENCE 1342 AA; 148097 MW; 7201E7F66CA37ABD CRC64;  
 Query Match 36.3%; Score 1435; DB 1; Length 1342;  
 Best Local Similarity 43.6%; Pred. No. 1.8e-101;  
 Matches 278; Conservative 92; Mismatches 244; Indels 24; Gaps 12;

QY 188 RSACHPSCPMKSGKRCWGESEDCOSLTRTYVCAGGC-ARCKGPIPTDCCHQCAAGCTG 246  
 DB 179 NGRSCPPCHEVKG-KCMGPGSEDCOTLTFTTCARQCNHCGRPNQCCHPCAGGCG 237  
 QY 247 PRHSDCLACLNHNHSGICELCPALVTYNTDTFEESMPNDEGRYTGASCVTACPNYVIST 306  
 DB 238 PDOTDCAFNRHNDGACVPRCPOLYVKNLFTOLEPNNHTKYQGVGVASCYNNFV-V 296  
 QY 307 DVGSCITVCPNLNOEYTAEDGTQRCBEKSKPCARVCYGLGMHLEKVRAYTSANIQERAG 366  
 DB 297 DQTSVCVRACPPDKMEVD-KNGLKMCPECGCLCPKACEGSGS--SRFQVDSNNIDGFVN 353  
 QY 367 CKRIGSLAFLEESPDGSPASNTAPLOPQLOVFETLEITGYLYISAMPDLPDLSYRQ 426  
 DB 354 CKRIKLNDELFTTGLNGDPNWKIRPALDPKLVNFTVREITTYLQINSPPHMHNFVSFS 413  
 QY 427 NLQVIRGLRHNGAYS-LTLQGLGISWGLRSRLREIGSGLALIHNTHLCEVHTVPMQDL 485  
 DB 414 NLTTIGGRSLYKRGSLIMKLNLYTSIGFBSLKEISAGRITYISANRQCLHNSLWTKV 473  
 QY 486 FNNPQALHTA-NRPEDECVGEGLACQOLCARGCMWGPPOCVNCSOFLNGQCEVEEC 544  
 DB 474 IKGPEERLDIKHNNBRRCDCVAEGKVCDPCLSSGGCMGPGQCLSCRMYSGGVCVTHC 533  
 QY 545 RVLOGLPREYVARHCLPCHPECPONGSVTCFGEADQCVACAHYKDPFCVAPCPGV 604  
 DB 534 NFLNDEPRFEHAECFCHPECPMEGTATNGSSSDTCAQCAHFRDGFHCYSSCPHGV 593  
 QY 605 KPDLSYPIWKPRDEGACQPCPCINCTHSCVDLDDKGC 642  
 DB 594 LG-AKGPYIKYRPDVQNECRPCHEMCTOGCKRPELDQC 629

## RESULT 10

ERB3\_RAT  
 ID 062799; 062955;  
 AC 15-DEC-1998 (Rel. 37, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)  
 DE (c-erbB3).  
 GN ERB3.  
 OS Rattus norvegicus (rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 NCBI\_TaxID=10116;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-Sprague-Dawley; TISSUE-Liver;  
 RA MEDLINE=96096535; PubMed=8522190;  
 RA Hellyer N.J., Kim H.-H., Greaves C.H., Sierke S.L., Roland J.G.;  
 RT "Cloning of the rat ErbB3 cDNA and characterization of the  
 RT recombinant protein.";  
 RL Gene 165:279-284(1995).  
 RN [2]  
 RP REVISIONS TO 85; 513 AND 565.  
 RA Hellyer N.J., Roland J.G.;  
 RL Submitted (Dec-2001) to the EMBL/Genbank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 922-1097 FROM N.A.  
 RC STRAIN-Sprague-Dawley; TISSUE-Sciatic nerve;  
 RA MEDLINE=97184212; PubMed=9030624;  
 RA Carroll S.L., Miller M.L., Frohneart P.W., Kim S.S., Corbett J.A.;  
 RT "Expression of neurotrophins and their putative receptors, ErbB2 and  
 RT ErbB3, is induced during Wallerian degeneration.";  
 RL J. Neurosci. 17:1642-1659(1997).  
 CC -I- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTRK.  
 CC -I- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 CC tyrosine phosphate.  
 CC -I- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS  
 CC (POTENTIAL).  
 CC -I- SUBCELLULAR LOCATION: Type I membrane protein.

CC -1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE  
 CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.  
 CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES  
 CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF  
 CC PHOSPHATIDYLINOSITOL 3-KINASE.  
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U09339; AAC38498.2; -  
 DR EMBL: U05330; AAC53050.1; -  
 DR HSSP: P11362; 1FGK.  
 DR InterPro: IPR000494; EGFR\_L\_domain.  
 DR InterPro: IPR000719; Euk\_kinase.  
 DR InterPro: IPR002174; Furin-like.  
 DR InterPro: IPR001245; Tyr\_kinase.  
 DR Pfam: PF00069; kinase.1.  
 DR Pfam: PF00757; Furin-like.1.  
 DR Pfam: PF01030; Recep\_L\_domain.2.  
 DR PRINTS: PR00109; TYRKINASE.  
 DR ProDom: PD000001; Euk\_pkinase.1.  
 DR SMART: SM00261; Fuv\_5.  
 DR SMART: SM00219; TyrcKc.1.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; FALSE\_NEG.  
 DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR; 1.  
 DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR Transmembrane: Glycoprotein; Multigene family; Receptor; Signal;  
 KW Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.  
 FT SIGNL 1 19  
 FT CHAIN 20 1339  
 FT DOMAIN 20 643  
 FT TRANSMEM 644 662  
 FT DOMAIN 663 1339  
 FT DOMAIN 183 259  
 FT NP\_BIND 707 964  
 FT BINDING 713 721  
 FT ACT\_SITE 740 740  
 FT ACT\_SITE 832 832  
 FT DISULFID 186 186  
 FT DISULFID 190 202  
 FT DISULFID 210 218  
 FT DISULFID 214 226  
 FT DISULFID 227 235  
 FT DISULFID 231 243  
 FT DISULFID 246 255  
 FT DISULFID 259 286  
 FT DISULFID 290 301  
 FT DISULFID 305 320  
 FT DISULFID 323 327  
 FT DISULFID 327 327  
 FT DISULFID 504 517  
 FT DISULFID 520 529  
 FT DISULFID 533 549  
 FT DISULFID 556 573  
 FT DISULFID 576 585  
 FT DISULFID 589 610  
 FT DISULFID 613 621  
 FT DISULFID 617 629  
 FT CARBOHYD 126 126  
 FT CARBOHYD 250 250  
 FT CARBOHYD 353 353  
 FT CARBOHYD 408 408  
 FT CARBOHYD 414 414  
 FT CARBOHYD 437 437  
 FT CARBOHYD 469 469  
 FT CARBOHYD 522 522  
 FT CARBOHYD 566 566

FT CARBOHYD 616 616 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 1028 1028 L -> P (IN REF. 3).  
 SQ SEQUENCE 1339 AA: 147545 MW: 0A5F2402BBDFD1E CAC64;  
 Query Match 34.8%; Score 1374.5; DB 1; Length 1339;  
 Best Local Similarity 42.6%; Pred. No. 7.4e-97;  
 Matches 272; Conservative 87; Mismatches 250; Indels 29; Gaps 13;  
 QY 3 LALACRGLALLALPPAA---SYOVCTGDMKRLRPASETHIDMLRHLYOGCOVVOGN 59  
 DB 7 LQVLC---FLSLARSEMGNSQAVCPGTLNGSLVGDADNOYQTLTKYCEVVMGN 62  
 QY 60 LELVLTPTNLSLELDIOIEOVGVLLAHNOVROVPLQRLRIYRGTLQFEDNYALVDN 119  
 DB 63 LELVLTGHNMDLSFLQWIREVTGYLVANNEFSVLPRLNRYVRGTYGVGKALIFM-- 120  
 QY 120 GDLPLNTPYTGASPGGLRELQRLSLTELLKGVLIQRPOLCYQDTILMKDIFHNKQL 179  
 DB 121 ---LNYNT---NSSHALROLKFTOLTRELLSGVYIEKNKRLCHMDTIDRDIYVR-- 170  
 QY 180 ALTLIDFNRSRACHPCSPMKGSRMGSESDOSLFRVYACAGGC-ARCKGPLPTDCHE 238  
 DB 171 GAELIVKNNAGANCPCHEVCKG-RCWGEGPDQQLTKTICAPQCNGRCGPNPNOCCHD 229  
 QY 239 QCAAGCTGPKHSDCLACIENHSICELHCPALVTYNTDFESMPNEGRYTFGASCVTA 298  
 DB 230 ECAGGCGSQDDTDFACRRRNDSGACVRCPEPLVYKLFQLEPNHTKYQGVGVAS 289  
 QY 299 CPYVLTSTDVSGCLVPLRNLQRYTADGTQRCCKGKPCPARCYGLGMEHLEVRVTS 358  
 DB 290 CPNHFV-VDQTFVCRACPDPKMEVD-RHGLKMEPCGGLCPKAGEGSG--SRQYVDS 345  
 QY 359 ANIOEFAGCKKIFGSLAFLEPSFDGDSANTAPLQVFEETLEITGYLYISAPDS 418  
 DB 346 SNIDGFNCKRIKGLNDFILTLGNVDPMKRIPALDEPKLVFRVRIITGLINQSMRP 405  
 QY 419 LPDLVSFQNTQVIRGRILHNGAVS-LTLQIGISWLGSLRLREGSLALIHNTLCFV 477  
 DB 406 MHNFSVSNLTIGGRSLYNGFSLIMKNLNTVSLGFSRLKETSAGRVYISANOQCYH 465  
 QY 478 HTVWVDLFFNPQALHTA-NRPEDCEVGEGLACHOLCARGHMGWGPQCVNCSQFLR 536  
 DB 466 HSLWTRLLRGPSEERLDIKYDRPLGELAGKVCDDLSSGCGWGPQCCJSCRNVSR 525  
 QY 537 GOECVEECRVLYGPREYVARHCLPCHPCOPQNSVTFEGPEADQVACAHKDDPPFC 596  
 DB 526 EGVVTHCNFLQGRPREFVIEAQCFSCHPEDCLPHEGISTCNGSSGSDACARCAHRDQPHC 585  
 QY 597 VAPCPGVPKPDLSYMPIWKFPEDEGACQPCPINCTHSC 634  
 DB 586 VNSCPBGILG-AKGPYKYKYPDAQNECRPCHECTQGC 621  
 RESULT 11  
 XMRK\_XIPMA STANDARD; PRT; 1167 AA.  
 ID XMRK\_XIPMA  
 AC P1338;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Melanoma receptor protein-tyrosine kinase precursor (EC 2.7.1.112).  
 GN XMRK OR TU.  
 OS Xiphophorus maculatus (Southern platyfish).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Altieriomorpha;  
 OC Cyprinodontiformes; Poeciliidae; Xiphophorus.  
 RN NCBI\_TaxID=8083;  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90015140; PubMed=2797166;  
 RA Witbrodt J., Adam D., Malitschek B., Maueier W., Raulf F.,  
 Telling A., Robertson S.M., Scharf M.,

RT "Novel putative receptor tyrosine kinase encoded by the melanoma-inducing tu locus in Xiphophorus.";  
 RL Nature 341:415-421(1989).  
 RN [2]  
 RP REVISION TO 515.  
 RA Schartzl M.;  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 CC - FUNCTION: PROBABLE RECEPTOR WITH PROTEIN-TYROSINE KINASE ACTIVITY.  
 CC - CATALYTIC ACTIVITY: ATP + a protein tyrosine -> ADP + protein tyrosine phosphate.  
 CC - SUBCELLULAR LOCATION: Type I membrane protein.  
 CC - DISEASE: INVOLVED IN PIGMENT CELLS MALIGNANT MELANOMAS.  
 CC - SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation-CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/announce/or\\_send\\_an\\_email\\_to\\_license@sib-sib.ch](http://www.isb-sib.ch/announce/or_send_an_email_to_license@sib-sib.ch)).  
 CC -----  
 DR EMBL, X16891, CA34770.2, -;  
 DR PIR, S06142, S06142.  
 DR HSSP, P11362, 1FGK.  
 DR Interpro: IPR000494; EGFR\_Ldomain.  
 DR Interpro: IPR000719; Euk\_pkinase.  
 DR Interpro: IPR002174; Furin-like.  
 DR Interpro: IPR002290; Ser\_thr\_pkinase.  
 DR Interpro: IPR001245; Tyr\_pkinase.  
 DR Pfam: PF00069; Pkinase; 1.  
 DR Pfam: PF00757; Furin-like; 1.  
 DR Pfam: PF01030; Recept\_Ldomain; 2.  
 DR PRINTS: PR00109; TYRKINASE.  
 DR ProDom: PD000001; Euk\_pkinase; 1.  
 DR SMART: SM00261; FU; 5.  
 DR SMART: SM00220; S\_TKc; 1.  
 DR SMART: SM00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR; 1.  
 DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
 KW Transmembrane; Glycoprotein; Receptor; Signal; Transferase;  
 KW Tyrosine-protein kinase; ATP-binding; Phosphorylation; Proto-oncogene.  
 FT SIGNAL 1 25  
 FT \* CHAIN 26 1167 MELANOMA RECEPTOR PROTEIN-TYROSINE KINASE.  
 FT DOMAIN 26 642 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 643 665 POTENTIAL.  
 FT DOMAIN 666 1167 CYTOPLASMIC (POTENTIAL).  
 FT NP\_BIND 710 977 PROTEIN KINASE.  
 FT BINDING 743 724 ATP (BY SIMILARITY).  
 FT ACT\_SITE 835 835 ATP (BY SIMILARITY).  
 FT DISULFID 195 204 BY SIMILARITY.  
 FT DISULFID 199 212 BY SIMILARITY.  
 FT DISULFID 220 228 BY SIMILARITY.  
 FT DISULFID 224 236 BY SIMILARITY.  
 FT DISULFID 237 245 BY SIMILARITY.  
 FT DISULFID 241 253 BY SIMILARITY.  
 FT DISULFID 256 265 BY SIMILARITY.  
 FT DISULFID 269 296 BY SIMILARITY.  
 FT DISULFID 300 311 BY SIMILARITY.  
 FT DISULFID 315 330 BY SIMILARITY.  
 FT DISULFID 333 337 BY SIMILARITY.  
 FT DISULFID 504 513 BY SIMILARITY.  
 FT DISULFID 508 521 BY SIMILARITY.  
 FT DISULFID 524 533 BY SIMILARITY.  
 FT DISULFID 537 553 BY SIMILARITY.  
 FT DISULFID 556 569 BY SIMILARITY.  
 FT DISULFID 560 577 BY SIMILARITY.  
 FT DISULFID 593 615 BY SIMILARITY.  
 FT DISULFID 618 626 BY SIMILARITY.  
 FT DISULFID 622 634 BY SIMILARITY.

FT CARBOHYD 114 114 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 144 144 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 201 201 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 356 356 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 365 365 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 398 398 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 501 501 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 621 621 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 1167 AA; 129934 MW; 4793E0749DC1D55A CRC64;

Query Match 33.6%; Score 1330; DB 1; Length 1167;  
 Best Local Similarity 41.7%; Pred. No. 1,6e-93;  
 Matches 269; Conservative 93; Mismatches 259; Indels 24; Gaps 11;

QY 4 AALCFMGLLALLPPGAAT---OYCTGTDMKRLRSPASETHLMDLRHLYOGCOVOGN 59  
 DB 8 AALLD--LLLVLSIRCSCTDPDRKVCQGTSMQMTM---LDNHVYLMKKMYSGCVNLEEN 62  
 QY 60 LEETLTPNASTFTLQDIEVGVYLIANQYRVPDRLRYRGTOLEFEDNYALAVDN 119  
 DB 63 LEITTYQENQDLSFLOSIOEVGVYLIANNEVSTIPLVNLRLIRQONLYEGNFTLVMSN 122  
 QY 120 GDPNNTPVTGASPGSLREQLRSITELTKGVLIQRNPOLCYODTILMKDIFPKNNOL 179  
 DB 123 YK--NPSSP--DYOVGLKQLOLSNITELSGVAVSHNPILCANVETIMWDIVKTSNP 179  
 QY 180 ALTLIDTNRSRACHPCSPCKSRGCVSESDQSLFTYVACAGG--ARCKGLPTDCHE 238  
 DB 180 TGNLIPHAVERQCKQCDHCVCVNGSCWAPRGHCQKFTKLCAEQGNRRCRGKPIDCCNE 239  
 QY 239 QCAAGCTGPKHSDCLACHFNHSGICEPLVYTYNDTFESMPNPERRTFFGASCTYA 298  
 DB 240 HCAGCGTGRATDCLACRPFNDGTCDCPKPIYDIYSQVYNPNIKYTFGAACYKE 299  
 QY 299 CPYNTLPTDVGSCITVPLPHNOEYTAEDGTORCEKSKCAVCYGLGMEHLREVRAYS 358  
 DB 300 CENSNVYVE--GACVASCASGMLFVD--ENGRSKCRKQDVCPRVCGIGIGSLSNITAVNS 357  
 QY 359 ANIOEFGACKIIFGSLAFIPESFDGDPASNTAPLOPEOLQVETLEETIYGLYISAMPDS 418  
 DB 358 TNRFSNSTCKINGDILINRNSFEGBDPHYKICTMPEHLMLNTYKETIGYLVIMWPPEN 417  
 QY 419 LPDLVFPQNLQYIRGRIHLNGAYS--LTQGLGISMLGLRSLBELSGLALIHNNHLCFV 477  
 DB 418 MSLSVFQNLLEIRGRITFSRGSFVVQVYRHLQMLGRSLKVESAGVILNLTQLRYA 477  
 QY 478 HTVPMDOLFRRNHOALLHTANRPEDCEVGEGLACHOLCARGHCWGPPTOCVNCQSOLRG 537  
 DB 478 NTIINRRLEFSDQSIETDART-----ENQTCNNEGSEDCMGWGPPTMVSCLHVDRG 530  
 QY 538 QCEVECRVYLOGLPREYVNAHRLCPHCECPONGSVYTCFGEADQCAVAAHYKDPPECV 597  
 DB 531 GRCVASCNLLQEPREDAQYDGRVCHOECVLQYDLSLTCYGRGPNCSAHRFGQPCI 590  
 QY 598 ARCPGSGVAPDLSYMPIMKFPDEEGACQPCPTICTSHCDYLDKGC 642  
 DB 591 PRCPHGILGDGDTL--IMKYADKMGQCPCHONCTGCGSGPLSGC 634

RESULT 12  
 EGF\_R\_DROME STANDARD; PRT: 1426 AA.  
 AC P04412; 061601; Q9W2G0; P81868;  
 DT 13-AUG-1987 (Rel. 05, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Epidermal growth factor receptor precursor (EG 2.7.1.112) (Egfr)  
 DE (Gurken receptor) (torpedo protein) (Drosophila relative of ERBB).  
 GN EGFR OR TOP OR C-ERBB OR DER OR CG10079.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;



OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS TYPES I AND II).  
 RX MEDLINE=94350209; PubMed=8070664;  
 RA Clifford R., Schupbach T.;  
 RT "Molecular analysis of the Drosophila EGF receptor homolog reveals  
 RT that several genetically defined classes of alleles cluster in  
 RT subdomains of the receptor protein.";  
 RL Genetics 137:531-550(1994).  
 RN [2]  
 RP REVISIONS.  
 RA Clifford R., Schupbach T.;  
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.  
 RX MEDLINE=85124611; PubMed=2982499;  
 RA Livneh E., Glazer L., Segal D., Schlessinger J., Shilo B.-Z.;  
 RT "The Drosophila EGF receptor gene homolog: conservation of both  
 RT hormone binding and kinase domains.";  
 RL Cell 40:599-607(1985).  
 RN [4]  
 RP SEQUENCE FROM N.A., CHARACTERIZATION, AND ALTERNATIVE SPLICING.  
 RC STRAIN=Oregon R.; TISSUE=Embryo;  
 RX MEDLINE=87002474; PubMed=3093080;  
 RA Schejter E.D., Segal D., Glazer L., Shilo B.-Z.;  
 RT "Alternative 5' exons and tissue-specific expression of the  
 RT Drosophila EGF receptor homolog transcripts.";  
 RL Cell 46:1091-1101(1986).  
 RN [5]  
 RP SEQUENCE FROM N.A. (ISOFORM TYPE I), TISSUE SPECIFICITY, AND MUTATION  
 RP ANALYSIS.  
 RX MEDLINE=99102120; PubMed=9882502;  
 RA Lesochin A.M., Yu S.-Y., Katz J., Baker N.E.;  
 RT "Several levels of EGF receptor signaling during photoreceptor  
 RT specification in wild-type, Ellipse, and null mutant Drosophila.";  
 RL Dev. Biol. 205:129-144(1999).  
 RN [6]  
 RP SEQUENCE FROM N.A. (ISOFORM TYPE I).  
 RC STRAIN=Berkely;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Ceuliker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Adamatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Mortan J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman J.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,  
 RA Burtis K.C., Busan D.A., Butler H., Chadiu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson R., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegyev J.,  
 RA Jalali M., Kalush F., Karpen G., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshireff A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Murry D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleby J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheefel F., Smith T.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Sytkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.-Y., Massarman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 RN [7]  
 RP SEQUENCE OF 959-1078 FROM N.A.  
 RC STRAIN=Daekwanryeong;  
 RX MEDLINE=85137938; PubMed=2983232;  
 RA Wadsworth S.C., Vincent W.S. III, Bilodeau-Wentworth D.;  
 RT "A Drosophila genomic sequence with homology to human epidermal  
 RT growth factor receptor.";  
 RL Nature 314:178-180(1985).  
 RN [8]  
 RP SEQUENCE OF 1133-1137; 1155-1167 AND 1209-1216 FROM N.A., AND MUTATION  
 RP ANALYSIS.  
 RX MEDLINE=92038942; PubMed=1936959;  
 RA Raz E., Schejter E.D., Shilo B.-Z.;  
 RT "Interallelic complementation among DER/Elb alleles: implications for  
 RT the mechanism of signal transduction by receptor-tyrosine kinases.";  
 RL Genetics 129:191-201(1991).  
 RN [9]  
 RP REVIEW.  
 RX MEDLINE=97248481; PubMed=9094709;  
 RA Perrimon N., Perkins L.A.;  
 RT "There must be 50 ways to rule the signal: the case of the Drosophila  
 RT EGF receptor.";  
 RL Cell 89:13-16(1997).  
 CC -1- FUNCTION: BINDS TO FOUR LIGANDS: SPITZ, GURKEN, VEIN AND ARGOS,  
 CC WHICH IS AN ANTAGONIST. TRANSDUCES THE SIGNAL THROUGH THE RAS-RAF-  
 CC MAPK PATHWAY. INVOLVED IN A TRIAD OF DEVELOPMENTAL DECISIONS.  
 CC CRITICAL FOR THE PROLIFERATION OF IMAGINAL TISSUES, AND FOR THE  
 CC DETERMINATION OF BOTH THE ANTERO-POSTERIOR AND DORSO-VENTRAL  
 CC POLARITIES OF THE OCYOTE. IN THE EMBRYO, PLAYS A ROLE IN THE  
 CC ESTABLISHMENT OF VENTRAL CELL FATES, MAINTENANCE OF AMNIOSERA  
 CC AND VENTRAL NEUROECTODERMAL CELLS, GERM BAND RETRACTION, CELL FATE  
 CC SPECIFICATION IN THE CENTRAL NERVOUS SYSTEM AND PRODUCTION OF  
 CC CUTICLE.  
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 CC tyrosine phosphate.  
 CC -1- SUBCELLULAR LOCATION: ISOFORMS TYPE I AND TYPE II ARE TYPE I  
 CC MEMBRANE PROTEINS. ISOFORM TYPE III IS A TYPE III MEMBRANE  
 CC PROTEIN.  
 CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS; TYPE I (SHOWN HERE), TYPE II AND  
 CC TYPE III; ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- TISSUE SPECIFICITY: UBQUITOUSLY EXPRESSED IN EMBRYOS. IN LARVAE,  
 CC UNIFORM EXPRESSION IS SEEN IN WING DISKS, GENITAL DISK, ANLAGEN OF  
 CC TESTIS AND OVARY, AND BRAIN CORTEX. IN EYE-ANTENNA DISK, HIGHEST  
 CC EXPRESSION IS ANTERIOR TO MORPHOGENETIC FURROW, LEVELS REMAIN HIGH  
 CC IN PHOTORECEPTOR PRECURSOR CELLS. THIS PATTERN IS REVERSED IN  
 CC POSTERIOR EYE DISK. IN ADULTS EXPRESSION IS HIGH IN BRAIN CORTEX  
 CC AND THORACIC AND ABDOMINAL GANGLIA.  
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AF052754; AAC08536.1; -;  
 CC EMBL: AF052753; AAC08536.1; JOINED.  
 CC EMBL: AF052754; AAC08535.1; -;  
 CC EMBL: AF052752; AAC08535.1; JOINED.  
 CC EMBL: K03054; AAA51462.1; -;  
 CC EMBL: K03417; AAA51460.1; -;  
 CC EMBL: K03416; AAA50965.1; -;  
 CC EMBL: K03418; AAA51461.1; -;  
 CC EMBL: AF109077; AAD26134.1; -;  
 CC DR



DR	EMBL:	Af109078:	AAD26132.1;	-	JOINED.
DR	EMBL:	Af109082:	AAD26132.1;	-	JOINED.
DR	EMBL:	Af109078:	AAD26133.1;	-	JOINED.
DR	EMBL:	Af109084:	AAD26133.1;	-	JOINED.
DR	EMBL:	Af109079:	AAD26130.1;	-	JOINED.
DR	EMBL:	Af109081:	AAD26130.1;	-	JOINED.
DR	EMBL:	Af109079:	AAD26131.1;	-	JOINED.
DR	EMBL:	Af109083:	AAD26131.1;	-	JOINED.
DR	EMBL:	Af109080:	AAD26135.1;	-	JOINED.
DR	EMBL:	AE003454:	AAE46732.1;	-	JOINED.
DR	EMBL:	X02293:	CAAB2157.1;	-	JOINED.
DR	EMBL:	X78920:	CAA55523.1;	-	JOINED.
DR	EMBL:	X78918:	CAA55521.1;	-	JOINED.
DR	EMBL:	X78919:	CAA55522.1;	-	JOINED.
DR	PfR:	A00640:	GOFEE.	-	JOINED.
DR	HSSP:	P13362:	IFGK.	-	JOINED.
DR	FlyBase:	FBgn0003731:	Egfr.	-	JOINED.
DR	InterPro:	IPR000496:	EGFR_L_domain.	-	JOINED.
DR	InterPro:	IPR000719:	Euk_pkinase.	-	JOINED.
DR	InterPro:	IPR001214:	Furin-like.	-	JOINED.
DR	InterPro:	IPR001245:	Tyr_pkinase.	-	JOINED.
DR	Pfam:	PF00069:	pkinase.1.	-	JOINED.
DR	Pfam:	PF00757:	Furin-like.1.	-	JOINED.
DR	Pfam:	PF01030:	Recep.L_domain.2.	-	JOINED.
DR	PRINTS:	PR00109:	TYRKINASE.	-	JOINED.
DR	ProDom:	PD000001:	Euk_pkinase.1.	-	JOINED.
DR	SMART:	SM00261:	FU_7.	-	JOINED.
DR	SMART:	SM00219:	TYRKc.1.	-	JOINED.
DR	PROSITE:	PS00107:	PROTEIN_KINASE_ATP.1.	-	JOINED.
DR	PROSITE:	PS00109:	PROTEIN_KINASE_TYR.1.	-	JOINED.
DR	PROSITE:	PS50011:	PROTEIN_KINASE_DOM.1.	-	JOINED.
KW	Transmembrane:	Glycoprotein:	Receptor:	-	JOINED.
KW	Transmembrane:	protein_kinase:	ATP-binding:	-	JOINED.
KW	Transmembrane:	signal:	Alternative splicing:	-	JOINED.
KW	Developmental:	protein.	-	-	JOINED.
FT	SIGNAL	1	30	-	POTENTIAL.
FT	CHAIN	31	1426	-	EPIDERMAL GROWTH FACTOR RECEPTOR.
FT	DOMAIN	31	868	-	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	869	889	-	POTENTIAL.
FT	DOMAIN	890	1426	-	CYTOSOLASMIC (POTENTIAL).
FT	DOMAIN	938	1198	-	PROTEIN KINASE.
FT	NP_BIND	944	952	-	ATP (BY SIMILARITY).
FT	BINDING	971	971	-	ATP (BY SIMILARITY).
FT	ACT_SITE	1063	1063	-	BY SIMILARITY.
FT	MOT_RES	902	902	-	PHOSPHORYLATION (BY PKC) (BY SIMILARITY)

Query Match	25.68;	Score 1013.5;	DB 1;	Length 1426;
Best Local Similarity	33.08;	Pred. No. 2.6e-69;		
Matches 231; Conservative	93;	Mismatches 280;	Indels 97;	Gaps 20;

Oy	24	QVCGTJDMKJLKPSPENHIDMLPHKVOGOVONNELTYLTP--NANSJLQDIOEYOG	82
Dp	100	KICIGTSRLSVSPENKEHNHRLNDRTRTCTYYDGNKJLWLPRENILDSFJONIREVYG	159
Oy	83	YVLIAHNOVOPJQJRLRYVGTOLE-----EDNYALVALDNGDPJLNNTPYVAGSPGL	137
Dp	160	YILISHDVKKVPPKJLOIJRGLTFGLSVEEKKYALFV-----TYSKM	203
Oy	138	RELDRSLTELKKGVLQORPOCYODTLIMKDIFKNNQJLALLTIDTRSRACHPCSP	197
Dp	204	YTLLEPDLRLDVLNQOVGFHNHYNLCHMRTJONSEIYVNGDAYYNDFTLPEBECKCHE	263
Oy	198	MCKSRRCMGSESDCCSLTRTVCAAGCA--RCKGRLPTDCCHQCAAGCTGPKHSDCLAC	255
Dp	264	SCTHG-CMGEGPKKQCFKSLKTSPOCAGAGRCVGRPKRECHHFCAGGCTGPPQKXCIAC	322
Oy	256	LHPNHSIGCEIHLCAIYUYNDFPESMPENEGRTFFCASCVTAPYUYSITDVGSCILAC	315
Dp	323	KNFDEAVSKRECCGPKMKYNTPTVLETLNDEGKAYATCVCRCP--CHILRDNGACVYRSC	381
Oy	316	PLHNOEYTAEDGTORCEKCSPPCARVOCYGLAMEHLREVPRAVTSANTIOEPACCKIFGSLA	375
Dp	382	PODKMDKGG-----CVPNGCPKCTCGVTYH-----AGNIDSFNCYIVAGNIR	428

```

0Y 376 FLPESEFDG--DPA$MTA-----PQBPQAOVAFETLEETIGYLXISAMPD$LDL$FON 427
Db 429 ILD0FFSEFQVAVNYMTGPRYIPDLPERRERFSTVKEITGLYIEGTHDP$RNL$YF$RN 488
0Y 428 L0VINGRLTHNGAV-SLTL0GLG$ISWGL$R$RELG$SLALIHNT$HLC$FVHTVPMQ0LF 486
Db 489 LETTHGR0L$MESMRPALAIYVSSLYSL$EMRNLK0I$SG$SVYI0HNRDLCYV$NIRMPAIQ 548
0Y 487 RNP$OALLHTANRPEDEVC$EGLC$HOLC$ARGH$GMPGPT0CVNCSOFLRQ0ECVE$C$RY 546
Db 549 KEPEKQVAVNENL$RADLC$EKNGTIC$SDQCN$DGC$MGATD0CLTK$FN$FN$GTICLADCGY 608
0Y 547 LOGLR$REYV$NARHCL$PCH$PECQ$OPNG$SVT$C$GP$EAD0C$VAC$AHYK$D$P$F$V$ARC$P$G$V$K 606
Db 609 ISNAVK--FDNRKTCIKH$E$CR-----TNGAGADH$Q0EC$V$HVRD$Q$H$C$V$SEC$PKN--- 657
0Y 607 DL$TYPI$TK$F$D$E$B$G$A$C$O$P$C$I$N$T$H$C$V$D$LD$D$K$G$C$P$A$O$R$A$S$P$L$T$S$O$N$E$D$L$P$A$S$P$L-- 664
Db 658 -----KYND-RGV$C$R$E$C$H$A$T$C-----DGC-----TGPKDTIGIGACTTC 690
0Y 665 -----D$T$F$Y$R$L$E$D$D$D$G$D$L$V$D$A$E$Y$U$P$O-$G$F$C$P 697
Db 691 NLA$IN$ND$AT$V$K$R$C$L$K$D$K$C$P$D$Y--FWEY$H$P$D$Q$S$L$K$P 730

```

RESULT 13	ID	ENTRY	STANDARD	PROT	AA
LT23_CAEEL	LT23_CAEEL				
AC	P24348;				
DT	01-MAR-1992 (Rel. 21, Created)				
DT	01-MAR-1992 (Rel. 21, last sequence update)				
DT	15-JUN-2002 (Rel. 41, last annotation update)				
DE	let-23 receptor protein-tyrosine kinase precursor (EC 2.7.1.112).				
GN	LET-23 OR KIN-7 OR ZK1067.1.				
OS	Caenorhabditis elegans.				
OC	Eukaryota; Metazoa; Nematoa; Chromadorea; Rhabditida; Rhabditoidae				
CC	Rhabditidae; Pelodierinae; Caenorhabditis.				
OX	NCBI_TaxID=6239;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	MEDLINE=91080919; PubMed=1979659;				
RA	Ariola R.V., Koga M., Mendel J.E., Ohshima Y., Sternberg P.W.;				
RT	"The let-23 gene necessary for Caenorhabditis elegans vulval induction encodes a tyrosine kinase of the EGF receptor subfamily."				
RL	Nature 348:693-699(1990).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=Bristol N2;				
RA	MEDLINE=96177760; PubMed=8604137;				
RT	Sakai T., Koga M., Ohshima Y.;				
RL	"Genomic structure and 5' regulatory regions of the let-23 gene in the nematode C. elegans."				
RN	J. Mol. Biol. 256:548-555(1996).				
RP	[3]				
RC	SEQUENCE FROM N.A.				
RA	STRAIN=Bristol N2;				
RT	Thomas K.;				
RL	Submitted (MAR-1996) to the EMBL/Genbank/DBJ databases.				
RN	[4]				
RP	MUTANTS.				
RC	MEDLINE=94147981; PubMed=8313880;				
RA	Ariola R.V., Les G.M., Sternberg P.W.;				
RT	"Mutations in the Caenorhabditis elegans let-23 EGFR-like gene define elements important for cell-type specificity and function."				
RL	EMBO J. 13:360-366(1994).				
RN	[5]				
RP	SUBCELLULAR LOCATION.				
RC	STRAIN=Bristol N2;				
RA	MEDLINE=99287744; PubMed=10359617;				
RT	Whitefield C.W., Bernard C., Barnes T., Hekimi S., Kim S.K.;				
RL	"Basolateral localization of the Caenorhabditis elegans epidermal growth factor receptor in epithelial cells by the PDZ protein LIN-10."				

RL Mol. Biol. Cell 10:2087-2100(1999).  
CC -1- FUNCTION: Tyrosine kinase receptor required for the induction of  
CC vulval differentiation. Possible receptor for the inductive signal  
CC required for vulval development. Activated by lin-3 and acts by  
CC way of let-60 Ras. The lin-3/let-23 pair is a simplified version  
CC of the mammalian neuroligin-ERBB network.  
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
CC tyrosine phosphate.  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Basolateral and  
CC apical membrane of cell junctions in epithelial vulval precursor  
CC cells.  
CC -1- TISSUE SPECIFICITY: Vulval precursor cells.  
CC -1- DEVELOPMENTAL STAGE: Expressed during L2 and L3 larval stages.  
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL: X57767; CAA0919.1; ALT-SEQ.  
DR EMBL: D63426; BAA09729.1; ALT-INIT.  
DR EMBL: Z70038; CAA93882.1; -  
DR PIR: S13422; S13422.  
DR HSSP: P11362; 1FGK.  
DR WormPep: zk1067.1; CE03840.  
DR InterPro: IPR000494; EGRF\_L\_domain.  
DR InterPro: IPR000719; Euk\_pkinase.  
DR InterPro: IPR002174; Furin-like.  
DR InterPro: IPR001245; Tyr\_pkinase.  
DR Pfam: PF00069; pkinase; 1.  
DR Pfam: PF00757; Furin-like; 1.  
DR Pfam: PF01030; Recep\_L\_domain; 2.  
DR ProDom: PD000001; Euk\_pkinase; 1.  
DR SMART: SM00261; FG; 6.  
DR SMART: SM00219; TYRKC; 1.  
DR ProSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR ProSITE: PS00109; PROTEIN\_KINASE\_TYR; 1.  
DR ProSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
DR Transfaser: Tyrosine-protein kinase; Developmental protein;  
KW ATP-binding; Phosphorylation; Transmembrane; Glycoprotein; Receptor;  
KW signal.  
FT SIGNAL. 1 22  
FT CHAIN 23 1333 LEFT-23 RECEPTOR PROTEIN-TYROSINE KINASE.  
FT DOMAIN 23 818 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 819 841 POTENTIAL.  
FT DOMAIN 842 1323 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 845 1152 PROTEIN KINASE.  
FT NP\_BIND 891 899 ATP (BY SIMILARITY).  
FT BINDING 919 919 ATP (BY SIMILARITY).  
FT ACT\_SITE 1010 1010 BY SIMILARITY.  
FT DISULFID 220 226 BY SIMILARITY.  
FT DISULFID 224 226 BY SIMILARITY.  
FT DISULFID 244 251 BY SIMILARITY.  
FT DISULFID 248 262 BY SIMILARITY.  
FT DISULFID 263 271 BY SIMILARITY.  
FT DISULFID 267 279 BY SIMILARITY.  
FT DISULFID 282 291 BY SIMILARITY.  
FT DISULFID 295 322 BY SIMILARITY.  
FT DISULFID 326 337 BY SIMILARITY.  
FT DISULFID 341 356 BY SIMILARITY.  
FT DISULFID 359 364 BY SIMILARITY.  
FT DISULFID 320 329 BY SIMILARITY.  
FT DISULFID 524 537 BY SIMILARITY.  
FT DISULFID 540 549 BY SIMILARITY.  
FT DISULFID 553 567 BY SIMILARITY.  
FT DISULFID 570 577 BY SIMILARITY.  
FT DISULFID 574 585 BY SIMILARITY.  
FT DISULFID 588 604 BY SIMILARITY.  
FT DISULFID 608 620 BY SIMILARITY.

FT DISULFID 623 632 BY SIMILARITY.  
FT DISULFID 627 644 BY SIMILARITY.  
FT DISULFID 647 660 POTENTIAL.  
FT DISULFID 670 703 POTENTIAL.  
FT DISULFID 696 703 POTENTIAL.  
FT DISULFID 700 715 POTENTIAL.  
FT DISULFID 717 731 POTENTIAL.  
FT DISULFID 735 750 POTENTIAL.  
FT DISULFID 753 763 POTENTIAL.  
FT DISULFID 757 771 POTENTIAL.  
FT DISULFID 774 787 POTENTIAL.  
FT DISULFID 791 805 POTENTIAL.  
FT CARBOHYD 91 91 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 169 169 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 255 255 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 376 376 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 561 561 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 635 635 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 746 746 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 776 776 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT VARIANT 368 368 C -> Y (IN SY10).  
FT VARIANT 469 469 C -> R (IN MN216).  
FT VARIANT 700 700 C -> W (IN MN23).  
FT VARIANT 733 733 C -> Y (IN SY11).  
FT VARIANT 1065 1065 T -> I (IN SY16).  
FT VARIANT 1074 1074 G -> E (IN SY7).  
SQ SEQUENCE 1323 AA; 150510 MW; 680307E53EEFA99 CRC64;

Query Match 18.2%; Score 718.5; DB 1; Length 1323;  
Best Local Similarity 28.0%; Pred. No. 7.2e-47;  
Matches 191; Conservative 107; Mismatches 256; Indels 129; Gaps 28;

QY 25 VCTGTGTMKRLPASPEHIDMLRHLYOGGOVGNELATYLPN----- 68  
DB 39 LCSGTNGISRYGTGNI -LEDLETMTRGCRRYVGNLEPIETIENETKKMRSTNVDR 97  
QY 69 -----ASLSFLDIOEVGYVLIANNOVROYELRLRYRGTLFEDNYALAVLDNGDP 122  
DB 98 NEDSPLKSIKINFNLBEIRGSLIIVANIGISFRLRYVYGEVPHDN -ALYIHRNDR 155  
QY 123 LANTTYTGASPGGLRELRLSITELLKGVLIQRNPOLCY -QDTILMKDIFKNNQAL 181  
DB 156 -----VHEVYRELRYIRNGSVTTIQDNPCKMYIDKIDMKELVDPD -VQ 199  
QY 182 TLIDTNRSRACH-----PCSPMKGSRGCMGSESEDCSLFTVYAGGACARC---KGPL 211  
DB 200 KVETTSHQHCYONGKSMKACHESC--NDKCGSGMDNDQRYVYVCPKSCSGCFYNSIS 258  
QY 232 PTDCHEQCAAGCTGPRHSDCLAFHNHSGICELCPALVYNTDTFESMPNDEGRYTF 291  
DB 259 SYCCDSACAGCGCTGHPKNCIACSKYELDGIICIECPKRIENHRTGRVLPNDRYQN 318  
QY 292 GASCYVACPNYL-STDVSGCLVLC-PLNQEVTAEDGVRCKC-SKRCARCYGLGHE 348  
DB 319 GNHCVKECPPELLIENDV--CYRHCSGDGHYDATTKD--VRECEKCRSSSPKICITYDG-- 372  
QY 349 HUREVAVTSANIOERGACCKKIGSLAFIPESPDGDPASVTAPLQEOLOVPEITBEITG 408  
DB 373 HL-----INETLKNIEGCGQIDGHL-IIEHA-----TEQLKLVETVAIVSE 414  
QY 409 YLYISAMPDLPDLVSFONLQVIRGILHNQAVSLTL-QGLGISWLGSLRSLRELASGLAL 467  
DB 415 YIITV--QQNFYDLKRLKNQIIEGRKLNHVRAALAIYQDDLEELSLMSIKIKGAVL 472  
QY 468 IHNHTLCFYHYVPWOLF-----RNPQALLHTANRPEDECGEGIACHQLARGCWGP 523  
DB 473 IMKNHRLCYVSKIDWSIITSKGDMPSLAIENRDSKICETEORVYCDKNCCKRCWCK 532  
QY 524 GPTQCVNCSQFLGSCVEECRVLOGLPREYVARBCLOCHPCOPQONSVCYCFGEADQ 583  
DB 533 EPEDCLECTKWSVGTCEVKCDI-KGFLNRQTSMK-CERCSPCE-----TCNGIGELD 584  
QY 584 CVACAH-----YKDPFCVARCPSGVKPDLSYMPIMKFPPEEGACQPCPINCISVD 636

```

DB 565 CLTCHKRLYNDSDFNRCHECHDC-----PVSHFPIQKRVCEKCHPTCY----- 628
OY 637 LDDKCPAEDQASPLTSONEDLG 659
DB 629 --DNCC-----TGPDSNLG 640

RESULT 14
ILPR_BRAVA STANDARD: PRT; 1363 AA.
ID ILPR_BRAVA
AC 002466;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Insulin-like peptide receptor precursor (EC 2.7.1.112) (ILP receptor).
OS Branchiostoma lanceolatum (Common lancelet) (Amphioxus).
OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
OC Branchiostoma.
OC NCBI_Taxid=7740;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96408719; PubMed=8813726;
RA Pashmforoush M., Chan S.J., Steiner D.F.;
RT "Structure and expression of the insulin-like peptide receptor from
RT amphioxus."
RL Mol. Endocrinol. 10:857-866(1996).
CC -1- FUNCTION: THIS RECEPTOR BINDS TO THE INSULIN RELATED PEPTIDE AND
CC HAS A TYROSINE-PROTEIN KINASE ACTIVITY.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: PROBABLE TETRAMER OF 2 ALPHA AND 2 BETA CHAINS LINKED BY
CC DISULFIDE BONDS. THE ALPHA CHAINS CONTRIBUTE TO THE FORMATION OF
CC THE LIGAND-BINDING DOMAIN, WHILE THE BETA CHAIN CARRY THE KINASE
CC DOMAIN (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).
CC -1- SIMILARITY: BELONGS TO THE TYR FAMILY OF PROTEIN KINASES. INSULIN
CC RECEPTOR SUBFAMILY.
CC -----
CC * THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; S83394; AAB50848.1; -.
DR HSSP; P06213; IIRK.
DR InterPro; IPR000494; EGFR_Ldomain.
DR InterPro; IPR000719; Euk_kinase.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR003962; FNIII_repeat.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR002011; RTKinaseI.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00041; fn3.3.
DR Pfam; PF00063; kinase.1.
DR Pfam; PF00757; Furin-like.1.
DR Pfam; PF01030; Recep_Ldomain.2.
DR PRINTS; PR00014; FNTYPEIII.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_kinase.1.
DR SMART; SM00060; FN3.3.
DR SMART; SM00261; FU; 1.
DR SMART; SM00219; TYRK; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP.1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR.1.
DR PROSITE; PS00239; RECEPTOR_TYR_KIN_III.1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM.1.
DR Transerfer; PS50011; PROTEIN_KINASE; Receptor; Transmembrane;
KW Glycoprotein; ATP-binding; Phosphorylation; Signal.
FT SIGNAL 1 29 POTENTIAL.

```

```

FT CHAIN 30 716 INSULIN-LIKE PEPTIDE RECEPTOR, ALPHA-
FT PROPEP 717 720 CHAIN (POTENTIAL).
FT CHAIN 721 1363 POTENTIAL.
FT DOMAIN 721 928 INSULIN-LIKE PEPTIDE RECEPTOR, BETA-
FT TRANSMEM 929 949 CHAIN (POTENTIAL).
FT DOMAIN 950 1363 EXTRACELLULAR (POTENTIAL).
FT NP_BIND 994 1283 POTENTIAL.
FT BINDING 1000 1008 CYTOPLASMIC (POTENTIAL).
FT ACT_SITE 1028 1028 PROTEIN KINASE.
FT MOD_RES 1174 1174 ATP (BY SIMILARITY).
FT CARBOHYD 51 51 BY SIMILARITY.
FT CARBOHYD 97 97 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 137 137 N-LINKED (GLCNAc...) (POTENTIAL).
FT CARBOHYD 278 278 N-LINKED (GLCNAc...) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAc...) (POTENTIAL).
FT CARBOHYD 599 599 N-LINKED (GLCNAc...) (POTENTIAL).
FT CARBOHYD 617 617 N-LINKED (GLCNAc...) (POTENTIAL).
FT CARBOHYD 665 665 N-LINKED (GLCNAc...) (POTENTIAL).
FT CARBOHYD 666 666 N-LINKED (GLCNAc...) (POTENTIAL).
FT CARBOHYD 711 711 N-LINKED (GLCNAc...) (POTENTIAL).
FT CARBOHYD 732 732 N-LINKED (GLCNAc...) (POTENTIAL).
FT CARBOHYD 736 736 N-LINKED (GLCNAc...) (POTENTIAL).
FT CARBOHYD 743 743 N-LINKED (GLCNAc...) (POTENTIAL).
FT CARBOHYD 816 816 N-LINKED (GLCNAc...) (POTENTIAL).
FT CARBOHYD 885 885 N-LINKED (GLCNAc...) (POTENTIAL).
FT CARBOHYD 898 898 N-LINKED (GLCNAc...) (POTENTIAL).
SQ SEQUENCE 1363 AA; 154104 MW; 238120B4EAB1ED65 CRC64;

```

Query Match 9.7%; Score 384; DB 1; Length 1363;  
 Best Local Similarity 26.3%; Pred. No. 2,3e-21;  
 Matches 137; Conservative 56; Mismatches 180; Indels 148; Gaps 25;

```

OY 9 WGLT-----LALLPFGASTGYCTGTDKRLPASPETHLDMLRLHYGCGVQVQGNLELT 63
DB 12 WAALLVLVIGLGLVNSNEEYICDSMDIRNR-----VSLRLD-ENCYIIEGYLDI- 61
OY 64 YLPTNASLFLDIOEVGYVLIHNOYRQV-----LQRLR-----IVRGTO 106
DB 62 -----LILDFEEDDYSGLAPNVEIYDFLVLRGLVLMSELPMIAVIRGTN 112
OY 107 LFEDNYVALVINDGNPLNNTPTVTGASPGGLRELDRLSTELKSGVLIQRNPOLCYODT 166
DB 113 LF-FNYALVFEMLD-----MOKIGLYSLQNTIRGVSRIEKNPNLCYIDT 156
OY 167 ILMKDIF---HKNNQLATLIDTNRSRAC-HPCSPMK-----GSRGCGESSEDCQSILT 216
DB 157 IDWSFIASGYSNM-----FIVDNREECVNCPCRCRIKRVYLDQLCMAE--EHGQKVC 210
OY 217 RTVCAGGCAKCKGRLPTDCHEQCAAGCTGPKHSCCLACLHFNHSGIELHCPALVYNT 276
DB 211 PESCIQNR---DGISCCHEHCIGCGDGPETRCVACKFVHNGECLICPPPTQYK 266
OY 277 D-----TFESMPBEGRY--TFGASCVTACPYVYSTDVGSCVLVPLINQVETADGQR 330
DB 267 DRCITTEECPTTNSVVKLHHRKCIPECPSTY-TTIDINPRL----- 308
OY 331 CEKSKPCARVCYGLGEMHLREVRVAVTSANIOEFAGCKKIGSLAFLPESFGDPAWYA 390
DB 309 CTECGQCPKSKGGLVSL-----AANQRFPGCIIIBELKISIRGDP----- 352
OY 391 PLQPEQLQVETLEETITYL-----YISAMPD-SLPDLSVFQNTQVIGRILHNGAYSLT 444
DB 353 -----NIIDLEENLGLIEVGHVAIVRSYALVTDLFLSKRIRGIQENG-YAFY 404
OY 445 LQGLGISTWLGRSLREL-----GSGLALIHNNHLC 475
DB 405 V-----LDNRNLEKLPDMQDRTITIDEGKLFFHFNPRLC 438
RESULT 15

```

IRR\_MOUSE  
ID IRR\_MOUSE STANDARD; PRT; 1300 AA.  
AC 09WTL4;  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Insulin receptor-related protein precursor (EC 2.7.1.112) (IRR)  
DE (IRR-related receptor).  
GN INSRR OR IRR.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99272238; PubMed=10342810;  
RA Hirayama I., Tamemoto H., Yokota H., Kubo S.-K., Wang J., Kuwano H.,  
Nagamachi Y., Takeuchi T., Izumi T.;  
RT "Insulin receptor-related receptor is expressed in pancreatic b-cells  
and stimulates tyrosine phosphorylation of insulin receptor  
substrate-1 and-2.";  
RT Diabetes 48:1237-1244(1999).  
RL  
CC -1- FUNCTION: This receptor probably binds an insulin related protein  
and has a tyrosine-protein kinase activity. It phosphorylates the  
insulin receptor substrates IRS-1 and IRS-2.  
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
tyrosine phosphate.  
CC -1- SUBUNIT: PROBABLE TETRAMER OF 2 ALPHA AND 2 BETA CHAINS LINKED BY  
DISULFIDE BONDS. THE ALPHA CHAINS CONTRIBUTE TO THE FORMATION OF  
THE LIGAND-BINDING DOMAIN, WHILE THE BETA CHAINS CARRY THE KINASE  
DOMAIN (By similarity).  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).  
CC -1- TISSUE SPECIFICITY: Highly expressed in the islets as well as in  
pancreatic beta-cells.  
CC -1- SIMILARITY: BELONGS TO THE TYR FAMILY OF PROTEIN KINASES. INSULIN  
RECEPTOR SUBFAMILY.  
CC -1- SIMILARITY: CONTAINS 3 FIBRONECTIN TYPE III-LIKE DOMAINS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See http://www.isb-sib.ch/announce/  
or send an email to license@isb-sib.ch).  
CC  
CC -----  
DR EMBL: AB007135, BAA7835.1, -  
DR HSSP: P06213, 1IRK.  
DR MGD; MG1:1346037, INSTR.  
DR InterPro: IPR000494; EGFR\_L.domain.  
DR InterPro: IPR000719; Euk\_pkinase.  
DR InterPro: IPR003961; FN\_III.  
DR InterPro: IPR002174; Furin-like.  
DR InterPro: IPR002011; RTKinaseII.  
DR InterPro: IPR001245; Tyr\_pkinase.  
DR Pfam: PF00041; fn3; 2.  
DR Pfam: PF00069; pkinase; 1.  
DR Pfam: PF00757; Furin-like; 1.  
DR Pfam: PF01030; Recep\_L.domain; 2.  
DR PRINTS: PR00109; TYRKINASE.  
DR PRODOM: PD000001; Euk\_pkinase; 1.  
DR SMART; SM00060; FN3; 3.  
DR SMART; SM00261; FU; 1.  
DR SMART; SM00219; TYRKC; 1.  
DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE: PS00109; PROTEIN\_KINASE\_DOM; 1.  
DR PROSITE: PS00239; RECEPTOR\_TYR\_KIN\_III; 1.  
DR Transferrase; Tyrosine-protein kinase; Receptor; Transmembrane;  
KW Glycoprotein; ATP-binding; Phosphorylation; Repeat; Signal.  
FT SIGNAL 1 26  
FT CHAIN 27 1300 INSULIN RECEPTOR-RELATED PROTEIN, ALPHA-  
FT CHAIN 27 746

FT CHAIN 747 1300 CHAIN (PROBABLE).  
FT TRANSMEM 747 921 INSULIN RECEPTOR-RELATED PROTEIN, BETA-  
FT DOMAIN 922 943 CHAIN (PROBABLE).  
FT DOMAIN 944 1300 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 944 1300 FIBRONECTIN TYPE-III 1.  
FT DOMAIN 602 811 FIBRONECTIN TYPE-III 2.  
FT DOMAIN 813 910 FIBRONECTIN TYPE-III 3.  
FT DOMAIN 979 1254 PROTEIN KINASE.  
FT NP\_BIND 985 993 ATP (BY SIMILARITY).  
FT BINDING 1013 1013 ATP (BY SIMILARITY).  
FT ACT\_SITE 1115 1115 BY SIMILARITY.  
FT DISULFID 214 222 BY SIMILARITY.  
FT DISULFID 216 228 BY SIMILARITY.  
FT DISULFID 229 237 BY SIMILARITY.  
FT DISULFID 233 246 BY SIMILARITY.  
FT DISULFID 249 258 BY SIMILARITY.  
FT DISULFID 262 274 BY SIMILARITY.  
FT DISULFID 280 300 BY SIMILARITY.  
FT DISULFID 304 317 BY SIMILARITY.  
FT DISULFID 320 324 BY SIMILARITY.  
FT DISULFID 657 864 INTERCHAIN (POTENTIAL).  
FT CARBOHYD 47 47 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT CARBOHYD 100 100 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT CARBOHYD 311 311 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT CARBOHYD 411 411 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT CARBOHYD 492 492 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT CARBOHYD 528 528 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT CARBOHYD 616 616 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT CARBOHYD 634 634 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT CARBOHYD 756 756 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT CARBOHYD 885 885 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT CARBOHYD 898 898 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT MOD\_RES 1145 1145 PHOSPHORYLATION (AUTO...) (POTENTIAL).  
SQ SEQUENCE 1300 AA; 144745 MW; 2989FC3C521A2067 CRC64;  
Query Match 9.7%; Score 383; DB 1; Length 1300;  
Best Local Similarity 28.4%; Pred. No. 2; 6e-21;  
Matches 152; Conservative 74; Mismatches 168; Indels 142; Gaps 34;  
OY 1 MELALICRWG--LLALLPQGA--STOVCTGDMKRLPASPETHLDMIRLYGCCQV 56  
1 MAAPALMPGWHVHLLMSLISGSLDLEVCPSLDIR-----SEVTELRRL-ENC5V 51  
DB  
OY 57 QGNLE--LTVLPNT--ASLSFLDIOEGVYLLAHNVQVPLRLR-----IVRG 104  
52 EGHQILLMFAATGEDEFGLSFPR-LTQVTDYLL--FRYGLSLRDLFPNLTIVRK 106  
OY 105 TQLEEDNYALAVLDNGDPLNNTPTVTCASPGLREQLRLSTFLKGVLIQHPOLCYO 164  
107 TRFL-LGYALILEMPH-----LRDVGPLSLGAVLNGAVKEKNOELCHL 150  
DB  
OY 165 DTLMKDFPKHKNQALTLTDTR--SRACHCSPKCK--S 202  
151 STIDW-----GLQAPAGTNNHVGKLGECADVCBPVLGAAGEPSCRTTSGRTDY 202  
OY 203 RCGESESDEQSLTRVCGACGACRKGRLPTDCDCHBCAGCTGPKH--SCLACLHFNHS 261  
203 RCM--TSSHCQKV--CPCPRGMACTAG--GDCCHSCLGCGQPEPRACVACRHLHYO 255  
DB  
OY 262 GICELHCPALVNTYNTDTFESMPNPEGRYTFGASCVA--CPY-----NYLSTDV-----GSC 311  
256 GVCRLACPP-----GYQYES-----CVTAELCAHLREVPGLATTFGIVEGSC 300  
DB  
OY 312 TLVCPHLNVDVLTADGTQREKSKPCARVCYGLGMEHLREVAAYVANSIQEAGCKKIF 371  
301 LAQCP--PGFTNRGSSIFCHKEGLOPKC-KVGRKTIQSVAT-----QDVLVGQTHVE 351  
DB  
OY 372 GSIAF-LPESFDDPASNTAPLOPELOVETLEETGTYLISAMPDSDLPSVFNLOY 430  
352 GNILTLNRQGN-----LEPELQRNLGLVETITGLFKR-HSPALVTLGFFRNKL 401

Mon Jan 13 17:52:22 2003

us-09-854-356-7.rsp

Page 21

Oy 431 IGRHLHNGAYSL-----TLGGLGISWL--GLRSLRELGSGLALIHNTLCPYH 478  
||| : : ||| : ||| ||| : ||| :  
Db 402 IRGSDWMDGWTFVLYLDNQNQLQGL-SWTAGL----TIVGKIFAFPRCLDEH 452

Search completed: January 13, 2003, 14:47:11  
Job time : 15.6039 secs

**THIS PAGE BLANK (uapfo)**